

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

Document AL7  
Appl. No. 09/517,466



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C12N 15/10, C07K 13/00	A1	(11) International Publication Number: WO 92/22650 (43) International Publication Date: 23 December 1992 (23.12.92)
--	----	--

(21) International Application Number: PCT/US92/04923

(22) International Filing Date: 12 June 1992 (12.06.92)

(30) Priority data:  
716,683 17 June 1991 (17.06.91) US

(71) Applicant: LIFE TECHNOLOGIES, INC. [US/US]; 8717  
Grovemont Circle, Gaithersburg, MD 20898 (US).

(72) Inventors: BEBEE, Robert, L. ; 442 N. Summit Avenue,  
#003, Gaithersburg, MD 20877 (US). HARTLEY,  
James, L. ; 7409 Hillside Drive, Frederick, MD 21703  
(US).

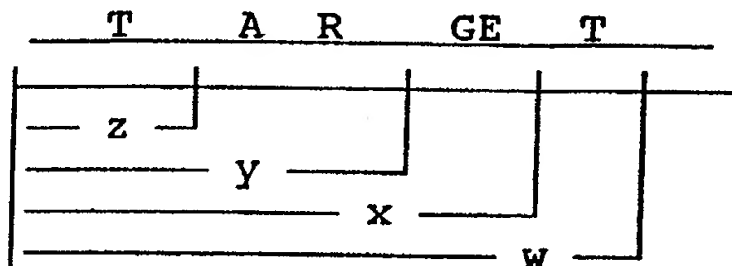
(74) Agents: AUERBACH, Jeffrey, I. et al.; Weil, Gotshal &  
Manges, 1615 L Street, N.W., Suite 700, Washington, DC  
20036 (US).

(81) Designated States: AT (European patent), BE (European patent), CA, CH (European patent), DE (European patent), DK (European patent), ES (European patent), FR (European patent), GB (European patent), GR (European patent), IT (European patent), JP, LU (European patent), MC (European patent), NL (European patent), SE (European patent).

Published  
With international search report.

(54) Title: RECOMBINATION-FACILITATED MULTIPLEX ANALYSIS OF DNA FRAGMENTS

	Size of bands
probe/primer I — <u>loxP</u> — CLON// — TARGET —	w
probe/primer I — <u>loxP</u> — CLON// — TARGE	x
probe/primer I — <u>loxP</u> — CLON// — TAR	y
probe/primer I — <u>loxP</u> — CLON// — T	z



Restriction sites in  
Target Molecule

(57) Abstract

The present invention contemplates the use of either generalized recombination, or more preferably, site-specific recombination to facilitate the sequence or fragment analysis of DNA molecules. Most preferably, the site-specific recombination system of bacteriophage P1 is employed to facilitate such sequence analysis.

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FI	Finland	ML	Mali
AU	Australia	FR	France	MN	Mongolia
BB	Barbados	GA	Gabon	MR	Mauritania
BE	Belgium	GB	United Kingdom	MW	Malawi
BF	Burkina Faso	GN	Guinea	NL	Netherlands
BG	Bulgaria	GR	Greece	NO	Norway
BJ	Benin	HU	Hungary	PL	Poland
BR	Brazil	IE	Ireland	RO	Romania
CA	Canada	IT	Italy	RU	Russian Federation
CF	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland	KR	Republic of Korea	SN	Senegal
CI	Côte d'Ivoire	LI	Liechtenstein	SU	Soviet Union
CM	Cameroon	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TG	Togo
DE	Germany	MC	Monaco	US	United States of America
DK	Denmark	MG	Madagascar		
ES	Spain				

-1-

TITLE OF THE INVENTION:RECOMBINATION-FACILITATED MULTIPLEX ANALYSIS OF  
DNA FRAGMENTSFIELD OF THE INVENTION:

5           The invention relates to the use of recombinases, and in particular, the Cre protein of bacteriophage P1 and its loxP DNA recombinational site, to facilitate the restriction fragment analysis and sequencing of a DNA molecule.

10   BACKGROUND OF THE INVENTION:

          The techniques of molecular biology were developed to analyze relatively small DNA molecules. Increasingly, however, research has centered on the analysis of larger and larger DNA molecules, such as the chromosomes of  
15   mammals, and in particular, the chromosomes of the human genome. The analysis of such large DNA molecules has often been limited by the ease with which the initially developed technology could be adapted to permit the analysis of such extremely large molecules. Such methods  
20   have exploited the ability of restriction endonucleases to produce fragments of the DNA molecule which would then be more amenable to sequence analysis.

-2-

# I. THE MAPPING OF RESTRICTION ENDONUCLEASE RECOGNITION SITES

One initial objective in the analysis of DNA molecules is to produce a gross physical map of the DNA. For small DNA molecules, this may be readily achieved using restriction endonucleases to identify and orient the corresponding recognition sites of such enzymes. Methods for performing such "restriction mapping" are well-known (see, for example, Perbal, B. A Practical Guide to Molecular Cloning, John Wiley & Sons, NY, (1984), pp. 208-216; Maniatis, T., et al. (In: Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Cold Spring Harbor, NY (1982), both herein incorporated by reference).

As will be apparent, the complexity of the data obtained in "restriction mapping" a target molecule increases rapidly as the size of the target molecule increases. For this reason, it is usually necessary to employ several strategies when attempting to obtain detailed maps of a large DNA molecule.

One strategy which is employed involves simultaneously digesting a target DNA molecule with combinations of several restriction enzymes each of which is expected to cleave at only a small number of sites in the target. For linear target molecules, the number of fragments equals the number of restriction enzyme sites plus 1. For a circular DNA molecule, the number of double-digestion fragments equals the number of fragments generated by the first enzyme plus the number of fragments generated by the second enzyme. Restriction maps are created from the data by a process which is part logic, and part trial and error (Lawn, R.M. et al., Cell 15:1157 (1978)).

The process of creating a restriction map may be facilitated by a sequential analysis of fragments. In this method, one treats a target molecule with a first restriction endonuclease, isolates the digestion products,

-3-

and then subjects the purified products to digestion with a second endonuclease. Such steps can be performed rapidly and efficiently (Parker, R.C. et al., Met. Enzymol. 65:358 (1980)).

5           In lieu of obtaining a complete endonuclease digestion of a target molecule, considerable information can be obtained by incubating the target molecule with endonucleases under conditions resulting in limited digestion. Two approaches have been developed. In the  
10 first, the aim is to compare the sizes of the partial- and complete-digestion products with one another, and to deduce which fragments might be adjacent to one another in the target molecule. In general, the number of partial-digestion products (F) of a linear DNA molecule that  
15 contains N+1 restriction sites (where N>0) is given by the formula:

$$F_{N+1} = \frac{N^2 + 3N}{2}$$

For a molecule having 20 sites, 209 partial-digestion products are obtainable. Thus, for large DNA molecules, the method is of very limited utility since the number of  
20 possible partial-digestion products quickly becomes unmanageable.

One means for simplifying such an analysis was proposed by Smith, H.O. et al., Nucleic Acids Res. 3:2387 (1976), herein incorporated by reference). This  
25 method uses a target molecule which has been labeled with <sup>32</sup>P at one of its termini. Digestion products are visualized by autoradiography after electrophoresis in agarose gels. Digestion products which are not linked to the labelled termini are not detected by the analysis.  
30 Thus, the number of labelled partial-digestion products is equal to the number of restriction sites within the target molecule. Moreover, the labelled fragments form a simple, overlapping ladder, with a common labelled terminus. The

-4-

order of ascension of the fragments corresponds to the order of restriction sites in the target molecule. Lastly, partial-digestion products produced through the action of several different enzymes can be analyzed simultaneously on the same gel.

One deficiency of the method is the difficulty which is often encountered in specifically labelling a single end of a target molecule. Indeed, due to the symmetrical nature of a double-stranded DNA molecule, both ends of the molecule are equally available for labelling. Thus, in practice, considerable difficulty is encountered in labelling only one end of a linear DNA molecule.

## II. THE SEQUENCING OF DNA MOLECULES

Initial attempts to determine the sequence of a DNA molecule were extensions of techniques which had been initially developed to permit the sequencing of RNA molecules (Sanger, F., J. Mol. Biol. 13:373 (1965); Brownlee, G.G. et al., J. Mol. Biol. 34:379 (1968)). Such methods involved the specific cleavage of DNA into smaller fragments by (1) enzymatic digestion (Robertson, H.D. et al., Nature New Biol. 241:38 (1973); Ziff, E.B. et al., Nature New Biol. 241:34 (1973)); (2) nearest neighbor analysis (Wu, R., et al., J. Mol. Biol. 57:491 (1971)), and (3) the "Wanderings Spot" method (Sanger, F., Proc. Natl. Acad. Sci. (U.S.A.) 70:1209 (1973)).

More recent advances in DNA sequencing have led to the development of two highly utilized methods for elucidating the sequence of a DNA molecule: the "Dideoxy-Mediated Chain Termination Method," also known as the "Sanger Method" (Sanger, F., et al., J. Mol. Biol. 94:441 (1975)) and the "Maxam-Gilbert Chemical Degradation Method" (Maxam, A.M., et al., Proc. Natl. Acad. Sci. (U.S.A.) 74:560 (1977), both references herein incorporated by reference).

-5-

A. DIDEOXY-MEDIATED CHAIN TERMINATION METHOD  
OF DNA SEQUENCING

In the dideoxy-mediated or "Sanger" chain termination method of DNA sequencing, the sequence of a DNA molecule is obtained through the extension of an oligonucleotide primer which is hybridized to the nucleic acid molecule being sequenced. In brief, four separate primer extension reactions are conducted. In each reaction, a DNA polymerase is added along with the four nucleotide triphosphates needed to polymerize DNA. Each of the reactions is carried out in the additional presence of a 2',3' dideoxy derivative of the A, T, C, or G nucleoside triphosphates. Such derivatives differ from conventional nucleotide triphosphates in that they lack a hydroxyl residue at the 3' position of deoxyribose. Thus, although they can be incorporated by a DNA polymerase into the newly synthesized primer extension, the absence of the 3' hydroxyl group causes them to be incapable of forming a phosphodiester bond with a succeeding nucleotide triphosphate. Thus, the incorporation of a dideoxy derivative results in the termination of the extension reaction. Since the dideoxy derivatives are present in lower concentrations than their corresponding, conventional nucleotide triphosphate analogs, the net result of each of the four reactions is to produce a set of nested oligonucleotides each of which is terminated by the particular dideoxy derivative used in the reaction. By subjecting the reaction products of each of the extension reactions to electrophoresis, it is possible to obtain a series of four "ladders." Since the position of each "rung" of the ladder is determined by the size of the molecule, and since such size is determined by the incorporation of the dideoxy derivative, the appearance and location of a particular "rung" can be readily translated into the sequence of the extended primer.

-6-

Thus, through an electrophoretic analysis, the sequence of the extended primer can be determined.

One deficiency of the dideoxy-mediated sequencing method is the need to optimize the ratio of dideoxy nucleoside triphosphates to conventional nucleoside triphosphates in the chain-extension/chain-termination reactions. Such adjustments are needed in order to maximize the amount of information which can be obtained from each primer. Additionally, the efficiency of dideoxy nucleotide incorporation in a particular target molecule is partially dependent upon the primary and secondary structures of the target.

The dideoxy-mediated method thus requires single-stranded templates, specific oligonucleotide primers, and high quality preparations of a DNA polymerase (typically the Klenow fragment of *E. coli* DNA polymerase I). Initially, these requirements delayed the wide spread use of the method. However, with the ready availability of synthetic primers, and the availability of bacteriophage M13 and phagemid vectors (Maniatis, T., et al., Molecular Cloning, a Laboratory Manual, 2nd Edition, Cold Spring Harbor Press, Cold Spring Harbor, New York (1989), herein incorporated by reference), the dideoxy-mediated chain termination method is now extensively employed.

#### 25 B. THE MAXAM-GILBERT METHOD OF DNA SEQUENCING

The Maxam-Gilbert method of DNA sequencing is a degradative method. In this procedure, a fragment of DNA is labeled at one end and partially cleaved in four separate chemical reactions, each of which is specific for cleaving the DNA molecule at a particular base (G or C) at a particular type of base (A/G, C/T, or A>C). As in the above-described dideoxy method, the effect of such reactions is to create a set of nested molecules whose lengths are determined by the locations of a particular base along the length of the DNA molecule being sequenced.



-7-

The nested reaction products are then resolved by electrophoresis, and the end-labeled molecules are detected, typically by autoradiography when a  $^{32}\text{P}$  label is employed. Four single lanes are typically required in order to  
5 determine the sequence.

The Maxam-Gilbert method thus uses simple chemical reagents which are readily available. Nevertheless, the dideoxy-mediated method has several advantages over the Maxam-Gilbert method. The Maxam-Gilbert method is  
10 extremely laborious and requires meticulous experimental technique. In contrast, the Sanger method may be employed on larger nucleic acid molecules.

Significantly, in the Maxam-Gilbert method the sequence is obtained from the original DNA molecule, and  
15 not from an enzymatic copy. For this reason, the method can be used to sequence synthetic oligonucleotides, and to analyze DNA modifications such as methylation, etc. It can also be used to study both DNA secondary structure and protein-DNA interactions. Indeed, it has been readily  
20 employed in the identification of the binding sites of DNA binding proteins.

Methods for sequencing DNA using either the dideoxy-mediated method or the Maxam-Gilbert method are widely known to those of ordinary skill in the art. Such methods  
25 are, for example, disclosed in Maniatis, T., et al., Molecular Cloning, a Laboratory Manual, 2nd Edition, Cold Spring Harbor Press, Cold Spring Harbor, New York (1989), and in Zyskind, J.W., et al., Recombinant DNA Laboratory Manual, Academic Press, Inc., New York (1988), both herein  
30 incorporated by reference.

## III. THE ANALYSIS OF LARGE DNA SEQUENCES

Both the above-described dideoxy-mediated method and the Maxam-Gilbert method of DNA sequencing require the prior isolation of the DNA molecule which is to be sequenced. The sequence information is obtained by subjecting the reaction products to electrophoretic analysis (typically using polyacrylamide gels). Thus, a sample is applied to a lane of a gel, and the various species of nested fragments are separated from one another by their migration velocity through the gel. The number of nested fragments which can be separated in a single lane is approximately 200-300 regardless of whether the Sanger or the Maxam-Gilbert method is used. Those of great skill in the art can separate up to 600 fragments in a single lane. Thus, in order to sequence large DNA molecules, it is necessary to fragment the molecule, and to sequence the fragments in separate lanes of the sequencing gel. The sequence of the entire molecule is obtained by orienting and ordering the sequence data obtained from each fragment.

Two approaches have been employed by those of skill in this art to accomplish this goal. In a random or shotgun sequencing approach, sequence data is collected by subcloning fragments of the target DNA molecule. No attempt is initially made to determine the linear orientation or order of the subclones with respect to the intact target DNA molecule. Instead, the accumulated data are stored and ultimately arranged into order by a computer (Staden, R., Nucleic Acids Res. 14:217 (1986); Anderson, S. et al., Nature 290:457 (1981); Gingeras, T.R., J. Biol. Chem. 257:13475 (1982); Sanger, F. et al., J. Mol. Biol. 162:729 (1982), and Baer, R. et al., Nature 310:207 (1984)). As will be appreciated, such random shotgun approaches often result in the multiple sequencing of the same oligonucleotide fragment, and thus are often inefficient in terms of time and materials.

-9-

In contrast, directed approaches have been employed in which sequences of the target DNA are obtained in a systematic fashion. For example, the target DNA molecule may be ordered by restriction mapping using the methods described above, and the discrete restriction fragments sequenced. Alternatively, the target molecule may be sequenced by sequencing nested sets of deletions which begin at one of its ends. The use of such nested fragments progressively brings more and more remote regions of the target DNA into range for sequencing. Lastly, sequence information obtained from a particular target molecule can be used to prepare a primer which can then be used in a subsequent sequencing reaction in order to obtain additional sequence information. As will be perceived, a directed sequence analysis of a target DNA molecule often requires substantial a priori information regarding the sequence. Moreover, for large target molecules (of sizes on the order of kilobases) such as would be encountered in the sequencing of eukaryotic (and in particular, mammalian) chromosomes, directional sequencing is quite arduous.

Several strategies have been developed to facilitate the sequence analysis of large (multi-kilobase) gene sequences. In one strategy, a large DNA molecule is fragmented through the use of restriction endonucleases which cut at infrequent sites. Such action results in the production of a small number of fragments each of which contains a portion of the sequence present in the original DNA molecule. Due to their smaller size, such fragments are more amenable to sequence analysis (using the above-stated methods) than the original DNA molecule. The sequence of the entire molecule is obtained by orienting the fragments with respect to each other in order to produce a gross physical map of the target molecule (Schwartz, D.C. et al., Cell 37:67-75 (1984); Southern, E.M. et al., Nucleic Acids Res. 15:5925-5943 (1987); Burke, D. et al., Science 236:808-812 (1987); Olson, M.V.

et al., Proc. Natl. Acad. Sci. (U.S.A.) 83:7826-7830 (1986)). Since this procedure reveals both the sequence and the orientation of the fragments, it permits one to readily determine the sequence of the entire DNA molecule.

5 Alternatively, a large DNA target may be subcloned into a large number of randomly selected bacteriophage or cosmid clones. Overlapping sequences in such clones are identified by unique restriction enzyme "fingerprinting" (Olson, M.V. et al., Proc. Natl. Acad. Sci. (U.S.A.)  
10 83:7826-7830 (1986); Coulson, A. et al., Proc. Natl. Acad. Sci. (U.S.A.) 83:7821-7825 (1986)). The information is then used to assemble a map of overlapping sets of clones (Staden, R., Nucleic Acids Res. 8:3673-3694 (1980)). This method has been successful in generating complete or par-  
15 tial maps of Saccharomyces cerevisiae chromosomes (Olson, M.V. et al., Proc. Natl. Acad. Sci. (U.S.A.) 83:7826-7830 (1986); C. Elegans (Coulson, A. et al., Proc. Natl. Acad. Sci. (U.S.A.) 83:7821-7825 (1986); Coulson, A. et al., Nature 335:184-186 (1988)); and the E. coli chromosome  
20 (Kohara, Y. et al., Cell 50:495-508 (1987)).

Several factors may limit the use of conventional methods in the analysis of the nucleotide sequence of a target molecule. Typically, each lane of a sequencing gel can resolve only about 300 different fragments. Thus, in  
25 order to determine the nucleotide sequence of a large DNA molecule, multiple sequencing gels are often needed. This, in turn, limits the amount of new sequence information which can be readily obtained per day. For a large nucleic acid molecule, a substantial number of  
30 technically demanding and time consuming steps must be performed. In particular, since the above-described techniques are capable of analyzing only one set of nested oligonucleotides per sample, the sequencing of large DNA molecules requires the use of multiple sequencing gels  
35 each having a large number of lanes. The electrophoretic analysis step in the sequencing process thus comprises a significant limitation to the amount of sequence

-11-

information which can be obtained and the rate with which it can be processed.

Similarly, the use of conventional methods in the analysis of restriction endonuclease-induced fragments of a target molecule is often not straightforward.

In some cases, sequence and restriction fragment analysis is limited by the low copy number of each target DNA sequence in natural genomes. One method for overcoming this limitation is through the use of amplification techniques, such as the polymerase chain reaction, "PCR" (Mullis, K. et al., Cold Spring Harbor Symp. Quant. Biol. 51:263-273 (1986); Erlich H. et al., EP 50,424; EP 84,796, EP 258,017, EP 237,362; Mullis, K., EP 201,184; Mullis K. et al., US 4,683,202; Erlich, H., US 4,582,788; and Saiki, R. et al., US 4,683,194), which references are incorporated herein by reference), the technique cannot readily be applied to amplify every target molecule present in a large gene sequence. A method for detecting and/or measuring PCR amplification is disclosed by Brenner, S. et al., in International Patent Application Publication No. WO/11375. This method entails linking a loxP sequence to a target molecule during PCR amplification. Amplification is detected by incubating the reacted molecules with Cre.

#### IV. MULTIPLEX ANALYSIS

A substantial improvement in DNA sequencing technology was recently developed, and designated "multiplex DNA sequencing" (Church, G.M., et al., Science 240:185-188 (1988); Church, G.M. et al., U.S. Patent 4,942,124; both herein incorporated by reference). Multiplex DNA sequencing utilizes DNA libraries which are individually constructed in 20 different plasmid vectors. In addition to standard drug resistance and replication origin elements, each of the vectors has a cloning site flanked by two different, predefined oligonucleotide

-12-

"tags" (i.e., forty total tags are used with twenty vectors). These tags are in turn flanked by sites recognized by the NotI restriction endonuclease (which cuts only at infrequent sites). The vectors differ from each other only by their tag sequences, which are originally selected from a random collection of chemically synthesized oligonucleotides.

In accordance with the method, DNA is sonicated to produce fragments of 900-1500 base pairs. Such DNA is rendered ligatable through treatment with Bal 31 exonuclease and then with T4 DNA polymerase and all four deoxynucleotide triphosphates. The DNA fragments are then ligated separately into each of the vectors and the ligation mixtures are used to transform E. coli cells. This procedure thus results in a formation of 20 gene libraries, which can then be amplified by conventional means. After amplification, the vectors are treated with NotI in order to excise the cloned DNA which is to be sequenced. Such excision produces DNA molecules having termini which are appropriate for the required subsequent chemical sequencing. The cloned DNA from each of the libraries is then mixed together to form a single pool containing each of the twenty members of the library.

The sequence of the cloned DNA of the libraries is determined using the Maxam-Gilbert method. The pool of 20 libraries is treated as a single unit in accordance with that method. The reaction products are then applied to a sequencing gel, and the oligonucleotides in the DNA sample are separated using gel electrophoresis. The DNA patterns, thus obtained, are then electro-transferred from the gels onto nylon membranes and crosslinked to the membranes using UV light.

Since each lane of the gel contains the reaction products of the sequencing of 20 different DNA molecules, each lane contains 20 overlaid ladders of sequence information. Because the NotI fragment of the cloned DNA contains the tag region of the vector, each oligonucleo-

-13-

5        tide of a particular sequence ladder contains the tag  
region. A particular sequence ladder may thus be  
visualized by hybridizing a labelled probe for a  
particular tag to the DNA bound to the membrane. By  
washing the membrane with sodium dodecyl sulfate and EDTA,  
it is possible to remove the hybridized probe from the  
membranes. This step thus prepares the membranes to be  
used to analyze a second sequence ladder by hybridizing a  
labelled probe for a second particular tag to the DNA  
10        bound to the membrane. In this manner, the sequence  
information of twenty vectors can be ascertained from a  
sequencing gel.

15        Thus, whereas conventional techniques permitted the  
sequencing of 300 bases per electrophoretic analysis, the  
multiplex DNA sequencing approach permits one to obtain  
sequence information of 6000 bases per analysis.

20        A significant advance in restriction endonuclease  
fragment analysis was recently disclosed by Garcia, E. et  
al. (In: Genome Mapping and Sequencing, Abstracts of  
Meeting Proceedings, Cold Spring Harbor, May 2-6, 1990,  
page 62). This reference concerns the use of a yeast  
artificial chromosome vector to clone large DNA sequences  
(such as from the human genome). To allow direct end-  
labelling of the vectors, the vectors were constructed to  
25        contain a 34 bp loxP fragment. By incubating this vector  
in the presence of Cre and a short labelled oligonucleo-  
tide which contained a loxP sequence, it was possible to  
label the molecules in vitro. The use of radioisotopic or  
biotin labels was disclosed.

30        Evans et al. have recently described a method which  
is potentially applicable to cloning, ordering clones, and  
the physical mapping of complex genomes (Evans, G.A. et  
al., Proc. Natl. Acad. Sci. (U.S.A.) 86:5030-5034 (1989),  
herein incorporated by reference). Unfortunately, Evans  
35        et al. have elected to refer to this method as "multiplex  
analysis." The term "multiplex analysis" as used herein  
differs significantly from the "multiplex analysis" term



-14-

used by Evans et al. Thus, although the Evans et al. reference uses the same term as that used by the inventors, it describes a different technique. The use of the term by the present inventors is consistent with its  
5 use by Church et al., and others in the art.

In brief, the method described by Evans et al. uses a single cosmid library which is constructed by inserting random DNA fragments into a site adjacent to a T3 or T7 bacteriophage promoters. Because these promoter sequences  
10 flank the cloned DNA (Wahl, G.N. et al., Proc. Natl. Acad. Sci. (U.S.A.) 84:2160-2164 (1987)), they can be used as probes to detect clones which have overlapping sequences.

In summary, a method which would minimize the number of gels needed for the determination of a particular  
15 sequence would, therefore, be highly desirable. Similarly, a method which would facilitate the construction of gross and fine restriction maps of a target molecule would also be highly desirable. Indeed, for the analysis of very large genomes, such as the human  
20 genome, the development of such methods may be essential.

#### SUMMARY OF THE INVENTION:

As indicated above, the analysis of a target DNA molecule often entails fragmenting the molecule, and analyzing and sequencing the resultant fragments.  
25 Especially for large DNA molecules, this is a difficult procedure. The present invention relates to an improved method for constructing gross and fine restriction maps of a target DNA molecule.

The invention further relates to an improved method  
30 for determining the nucleotide sequence of a target DNA molecule.

In detail, the invention provides a method for analyzing a target DNA molecule, which comprises:

(A) forming a recombinant molecule, the recombinant  
35 molecule comprising a probe/primer sequence linked to a



-15-

recombinational site (I), wherein the site is linked to the sequence of the target molecule;

(B) analyzing the target molecule using a nucleic acid molecule capable of hybridizing to the probe/primer sequence, or its complement.

The invention also provides the embodiment of the above method for determining nucleotide sequence of a target DNA molecule wherein the recombinant molecule is formed by:

(1) introducing the target DNA molecule into at least one vector, having a recombinational site (II), to thereby form a vector-target DNA construct;

(2) incubating the vector-DNA construct in the presence of a recombinase, and a DNA molecule having the recombinational site (I) and a probe/primer region; wherein the incubation is under conditions sufficient to permit the recombinase to mediate recombination between the recombinational site (II) of the vector-target DNA construct and the recombinational site (I) of the DNA molecule; and

(3) permitting the recombinase to mediate recombination between the recombinational sites, and to thereby form the sequencing molecule.

The invention also provides the embodiments of the above methods for determining nucleotide sequence of a target DNA molecule wherein at least two vector-target DNA constructs are formed, and wherein at least two different DNA molecules each having a recombinational site (I) and further having a different probe/primer region are employed; and wherein the determining of the sequence of the target molecule is through use of two probes, each capable of hybridizing to only one of the probe/primer regions, or its complement.

The invention also provides the embodiment of the above methods for determining nucleotide sequence of a target DNA molecule wherein the recombination is site-specific recombination, and, in particular, wherein in the

-16-

site-specific recombination, the recombinase is Cre, and at least one, and preferably both, of the recombinational sites (I) or (II) are loxP sites. The invention also provides the embodiments of the above methods wherein the recombinational site (I) is a loxP site, or a mutant loxP site, and wherein the vector contains one wild-type loxP site and one mutant loxP site.

The invention also provides the embodiment of the above-described method for analyzing a target DNA molecule, wherein the analysis comprises ordering restriction endonuclease recognition sites in a target DNA molecule, and wherein, in step (B), the analysis comprises (i) incubating the recombinant molecule in the presence of a restriction endonuclease under conditions sufficient to permit the endonuclease to cleave DNA containing a cleavage site recognized by the endonuclease; and

(ii) determining the order of any restriction sites in the target molecule using a nucleic acid molecule capable of hybridizing to a probe/primer sequence, or its complement.

The invention also provides the embodiment of the above method for ordering restriction endonuclease recognition sites in a target DNA molecule wherein the recombinant molecule is formed by:

- (1) introducing the target DNA molecule into at least one vector, having a recombinational site (II), to thereby form a vector-target DNA construct;
- (2) incubating the vector-DNA construct in the presence of a recombinase, and a DNA molecule having the recombinational site (I) and a probe/primer region; wherein the incubation is under conditions sufficient to permit the recombinase to mediate recombination between the recombinational site (I) of the DNA molecule, and the recombinational site (II) of the vector-target DNA construct;

-17-

(3) permitting the recombinase to mediate recombination between the recombinational sites (I) and (II), and to thereby form the recombinant molecule.

5 The invention also provides the embodiment of the above method for ordering restriction endonuclease recognition sites in a target DNA molecule wherein at least two vector-target DNA constructs are formed, and wherein at least two different DNA molecules each having a recombinational site (I) and further having a different  
10 probe/primer region are employed; and wherein the ordering of restriction sites of the target molecule is through use of two probes, each capable of hybridizing to only one of the probe/primer regions, or its complement.

The invention also provides the embodiments of the  
15 above methods for ordering restriction endonuclease recognition sites in a target DNA molecule wherein the recombination is site-specific recombination, and, in particular, wherein in the site-specific recombination, the recombinase is Cre, and at least one, and preferably  
20 both, of the recombinational sites (I) or (II) are loxP sites. The invention also provides the embodiments of the above methods for ordering restriction endonuclease recognition sites in a target DNA molecule wherein the recombinational site (I) is a loxP site, or a mutant loxP  
25 site, and wherein the vector contains one wild-type loxP site and one mutant loxP site.

The invention also provides a kit specially adapted to mediate recombination between a DNA molecule having a recombinational site (I), and a DNA vector, having a  
30 recombinational site (II), the kit comprising in close compartmentalization:

- 1) a first container containing a recombinase capable of mediating the recombination between the site (I) of the DNA molecule  
35 and the site (II) of the vector; and

-18-

- 2) a second container containing a DNA molecule having the recombinational site (I).

5 The invention also provides the embodiment of the above kit wherein the recombinase is Cre, and wherein at least one, and preferably both, of the recombinational sites (I) and (II) is a loxP site.

10 The invention also provides the embodiment of the above kit wherein the kit additionally contains a third container containing the DNA vector.

The invention also provides the embodiment of the above kits wherein the vector contains one loxP site, or wherein the vector contains one wild-type loxP site and one mutant loxP site.

15 The invention also provides a set of nested oligonucleotides each of which has a first region of unknown sequence, and a second region of known sequence, wherein the second region comprises both a recombinational site, and a probe/primer region.

20 The invention also provides the embodiments of the above set of nested oligonucleotides wherein at least one of the oligonucleotides is hybridized to a probe or to a primer.

25 The invention also provides the embodiment of the above set of nested oligonucleotides wherein the recombinational site is a loxP site.

#### BRIEF DESCRIPTION OF THE FIGURES

Figure 1 shows the recombination of a circular DNA molecule having two loxP sites in direct orientation.

30 Figure 2 shows the recombination of a circular DNA molecule having a single loxP site with a linear loxP-containing DNA molecule to produce a linear molecule

Figure 3 shows the recombination of the loxP sites in an inverted repeat orientation.

Figure 4 illustrates the exchange of the DNA that results from recombination between two linear molecules.

Figure 5 shows the use of Cre and loxP sites.

5 Figure 6 shows the structure of a vector in which the recombinational site is illustrated as a loxP site; the orientation of the loxP is always left to right relative to the other elements shown.

Figure 7 shows an alternative vector containing two non-corresponding recombinational sites.

10 Figure 8 shows the cloning of a target molecule into a vector.

Figure 9 shows the cloning of a target molecule into an alternative vector containing two non-corresponding recombinational sites.

15 Figure 10 shows the structures of loxP-containing oligonucleotide having one or more probe primer regions, wherein the roman numerals indicate the presence of the probe/primer regions which may be different or the same in sequence.

20 Figure 11 shows the linear molecule that is produced by incubating the molecules of Figures 8 and 9 together, in the presence of Cre.

Figure 12 shows the set of nested fragments that is obtained when the molecule of Figure 10 is subjected to partial restriction endonuclease digestion with a restriction enzymes, and analyzed by electrophoresis.

25 Figure 13 shows the visualization of the nested fragments, and how such visualization facilitates restriction mapping.

30 Figure 14 shows the linear molecule that result from recombination of the vector through the action of a suitable recombinase.

Figure 15 shows the structures of members of an unfractionated vector library after recombination with one of a plurality of linear molecules each of which differs from the other in the sequence of its probe/primer sequence.

-20-

Figure 16 shows the structures of loxP511-containing oligonucleotide having one or more probe primer regions, wherein the roman numerals indicate the presence of the probe/primer regions which may be different or the same in sequence.

Figure 17 shows the structure of a linear molecule containing more than one adjacent probe/primer regions separated by a recombinational site.

Figures 18A and 18B show the linear molecule that would be produced using the loxP / Cre system, after Cre-mediated recombination with a plasmid of the type shown in Figure 8 and with an Oligonucleotide I (Primer #n - loxP) or an Oligonucleotide II (Primer - Probe #n - loxP).

Figure 19 shows the structure of a cosmid containing recombinational sites.

Figure 20 shows a preferred, single-stranded loxP-containing oligonucleotide that possesses a sequence which causes it to snap back upon itself.

Figure 21 shows the result of recombination between the molecules of Figure 18 and Figure 19.

Figure 22 shows the structures of the molecules of an array of molecules obtained upon partial restriction endonuclease digestion of the molecules of Figure 21.

Figure 23 shows the structure of a class of molecules having two loxP sites in a direct repeat that result from the incubation of the mixture of oligonucleotides of Figure 22 with a DNA ligase in the presence of a second loxP-containing oligonucleotide, such as shown in Figure 20.

Figure 24 shows the structure of a class of molecules having two loxP sites in an inverted repeat that result from the incubation of the mixture of oligonucleotides of Figure 22 with a DNA ligase in the presence of a second loxP-containing oligonucleotide, such as shown in Figure 20.

Figure 25 shows a partially duplex molecule composed of certain oligonucleotides.

-21-

Figure 26 shows a partially duplex molecule composed of certain oligonucleotides.

DESCRIPTION OF THE PREFERRED EMBODIMENTS:

I. RECOMBINATION

5       The present invention uses the process of recombination (Watson, J.D., In: Molecular Biology of the Gene, 4th Ed., W.A. Benjamin, Inc., Menlo Park, CA (1987), which reference is incorporated herein by reference). Thus, an understanding of the process of recombination is desirable  
10       in order to fully appreciate the present invention.

      Recombination is a well-studied natural process which results in the scission of two nucleic acid molecules having identical or substantially similar sequences (i.e. "homologous"), and the joining of the two molecules such  
15       that one region of each initially present molecule becomes joined to a region of the other initially present molecule (Sedivy, J.M., Bio-Technol. 6:1192-1196 (1988), which reference is incorporated herein by reference). The recombinational reaction is catalyzed by enzymes, globally  
20       referred to as "recombinases." Such enzymes are naturally present in both prokaryotic and eukaryotic cells (Smith, G.R., In: Lambda II, (Hendrix, R. et al., Eds.), Cold Spring Harbor Press, Cold Spring Harbor, NY, pp. 175-209 (1983), herein incorporated by reference)). As discussed  
25       below, several recombinases are commercially available.

      Two types of recombinational reactions have been identified. In the first type of reaction, "general" or "homologous" recombination, any two homologous sequences can be recognized by the recombinase (i.e. a "general recombinase"), and thus act as substrates for the  
30       reaction. In contrast, the second type of recombination, "site-specific" recombination, employs specialized recombinases ( i.e. "site-specific recombinases") which can recognize only certain defined sequences. Thus, in



-22-

site-specific recombination, only molecules having a particular sequence may act as substrates for the reaction. The significance of each type of recombinational reaction is discussed below.

5

## A. GENERAL RECOMBINATION

General recombination is a process by which a "region" of DNA can be transferred from one DNA molecule to another. As used herein, a "region" of DNA is intended to generally refer to any nucleic acid molecule. The region may be of any length from a single base to a substantial fragment of a chromosome.

For general recombination to occur between two DNA molecules, the molecules must possess a "region of homology" with respect to one another. Such a region of homology must be at least two base pairs long. Two DNA molecules possess such a "region of homology" when one contains a region whose sequence is so similar to a region in the second molecule that homologous recombination can occur. The transfer of a region of DNA may be envisioned as occurring through a multi-step process.

If either of the two participant molecules is a circular molecule, then the above recombination event results in the integration of the circular molecule into the other participant.

The frequency of recombination between two DNA molecules may be enhanced by treating the introduced DNA with agents which stimulate recombination. Examples of such agents include trimethylpsoralen, UV light, etc.

The most characterized general recombination system is that of the bacterium E. coli (Smith, G.R., In: Lambda II, (Hendrix, R. et al., Eds.), Cold Spring Harbor Press, Cold Spring Harbor, NY, pp. 175-209 (1983)). The E. coli system involves the protein, RecA, which in the presence of ATP or another energy source, can catalyze the pairing



-23-

of DNA molecules at regions of homology. The RecA protein is commercially available from Pharmacia.

#### B. SITE-SPECIFIC RECOMBINATION

The above-described process of homologous recombination can occur between any two homologous DNA sequences. As indicated above, site-specific recombination can occur only between certain highly specialized and defined sequences. Site specific recombination is mediated between two such sequences through the action of one or more specialized enzymes. A large number of such site-specific recombination systems have been described. In particular, the P1, Flp, Gin/Fis, or  $\lambda$  recombinational systems may be employed. For the purposes of the present invention, the P1 site-specific recombinational system is preferred.

##### 1. The P1 Site-Specific Recombination System

A preferred site specific recombination system is that of the E. coli bacteriophage P1. Like bacteriophage  $\lambda$ , the P1 bacteriophage cycles between a quiescent, lysogenic state and an active, lytic state. The bacteriophage's site-specific recombination system catalyzes the circularization of P1 DNA (approximately 100 kb) upon its entry into a host cell. It is also involved in the resolution of multimeric P1 DNA molecules which may form as a result of replication or homologous recombination.

The P1 site-specific recombination system catalyzes recombination between specialized sequences, known as "loxP" sequences. The loxP site has been shown to consist of a double-stranded 34 bp sequence. This sequence contains two 13 bp inverted repeat sequences which are separated from one another by an 8 bp spacer region (Hoess, R., et al., Proc. Natl. Acad. Sci. (U.S.A.)

-24-

79:3398-3402 (1982); Sauer, B.L., U.S. Patent No. 4,959,317, herein incorporated by reference).

5 The recombination is mediated by a P1-encoded protein known as "Cre" (Hamilton, D.L., et al., J. Mol. Biol. 178:481-486 (1984), herein incorporated by reference). The Cre protein mediates recombination between two loxP sequences (Sternberg, N., et al., Cold Spring Harbor Symp. Quant. Biol. 45:297-309 (1981)). These sequences may be present on the same DNA molecule, or they may be present  
10 on different molecules. Cre protein has a molecular weight of 38,000. The protein has been purified to homogeneity, and its reaction with the loxP site has been extensively characterized (Abremski, K., et al., J. Mol. Biol. 259:1509-1514 (1984), herein incorporated by  
15 reference). The cre gene (which encodes the Cre protein) has been cloned (Abremski, K., et al., Cell 32:1301-1311 (1983), herein incorporated by reference). Cre protein can be obtained commercially from New England Nuclear/Dupont.

20 The site specific recombination catalyzed by the action of Cre protein on two loxP sites is dependent only upon the presence of the above-described thirty-four base pair loxP site and Cre. Magnesium ions or spermidine are needed for efficient recombination. Energy, however, is  
25 not required for this reaction; thus, there is no requirement for ATP or other similar high energy molecules. No proteins other than Cre are required in order to mediate site specific recombination at loxP sites (Abremski, K., et al., J. Mol. Biol. 259:1509-1514  
30 (1984)). In vitro, the reaction is highly efficient; Cre is able to convert up to about 70% of the DNA substrate into products and it appears to act in a stoichiometric manner. The extent of reaction reflects an equilibrium among the various molecules containing loxP sites.

35 Cre-mediated recombination can occur between loxP sites which are both present on the same molecule, or which are present on two different molecules. Because the

-25-

internal spacer sequence of the loxP site is asymmetrical, two loxP sites can exhibit directionality relative to one another (Hoess, R.H., et al., Proc. Natl. Acad. Sci. (U.S.A.) 81:1026-1029 (1984)). When two sites on the same DNA molecule are in a directly repeated orientation (i.e.  $\Rightarrow\Rightarrow\Rightarrow$ ), Cre will excise the DNA between the sites (Abremski, K., et al., Cell 32:1301-1311 (1983)). However, if the sites are inverted with respect to each other (i.e.  $\Rightarrow\Leftarrow\Leftarrow$ ), the DNA between them is not excised after recombination but is simply inverted. Thus, a circular DNA molecule having two loxP sites in direct orientation will recombine to produce two smaller circles, whereas circular molecules having two loxP sites in an inverted orientation simply invert the DNA sequences flanked by the loxP sites (Figure 1).

Two circular molecules each having a single loxP site will recombine to form a mixture of monomer, dimer, trimer, etc. circles. Higher concentrations of circles favor higher n-mers; lower concentrations of circles favor monomers.

A circular DNA molecule having a single loxP site will recombine with a linear loxP-containing DNA molecule to produce a linear molecule (Figure 2).

As indicated above, a linear molecule with direct repeats of loxP sites interacts to produce a circle (containing the sequences between the loxP sites), and a linear molecule. However, if the loxP sites are inverted repeats, recombination flips the sequence between the loxP sites back and forth (Figure 3).

When the starting DNA substrate is supercoiled, the final reaction product is also supercoiled (Abremski, K., et al., Cell 32:1301-1311 (1983); (Abremski, K., et al., J. Biol. Chem. 261:391-396 (1986)). The recombinational event does not, however, require supercoiling, and works with equal efficiency on supercoiled or linear molecules (Abremski, K., et al., Cell 32:1301-1311 (1983); Abremski, K., et al., J. Biol. Chem. 261:391-396 (1986)).

The nature of the interaction between Cre and a loxP site has been extensively studied (Hoess, R.P., et al., Cold Sprg. Harb. Symp. Quant. Biol. 49:761-768 (1984), herein incorporated by reference). In particular, mutations have been produced both in Cre, and in the loxP site.

The Cre mutants thus far identified have been found to catalyze recombination at a slower rate than that of the wild-type Cre protein. loxP mutants have been identified which recombine at lower efficiency than the wild-type site (Abremski, K., et al., J. Biol. Chem. 261:391-396 (1986); Abremski, K., et al., J. Mol. Biol. 202:59-66 (1988), herein incorporated by reference).

Of particular interest to the present invention is the loxP511 mutant site. The sequence of loxP511 is described by Hoess, R.H. et al. (Nucleic Acids Res. 14:2287-2300 (1986), herein incorporated by reference). Cre can mediate efficient recombination between two loxP sites, or between two loxP511 sites; it is, however, substantially incapable of mediating efficient recombination between a loxP site and a loxP511 site.

The Cre protein is capable of mediating loxP-specific recombination in *Saccharomyces cerevisiae* (Sauer, B., Molec. Cell. Biol. 7:2087-2096 (1987); Sauer, B.L., U.S. Patent No. 4,959,317, herein incorporated by reference). Such a property indicates that the Cre protein is capable of accessing DNA in eukaryotic cells even though such DNA is typically organized into nucleosomes within the nucleus, and bound to histones and other proteins.

Significantly, the loxP-Cre system can mediate site-specific recombination between loxP sites separated by extremely large numbers of nucleotides (Sternberg, N. (Proc. Natl. Acad. Sci. (U.S.A.) 87:103-107 (1990), herein incorporated by reference). Indeed, the ability of Cre to circularize the bacteriophage P1 evidences its ability to mediate the recombination of large DNA molecules. Recombination has been demonstrated to occur between two

-27-

loxP sites present on the 150 kb genome of the pseudo-rabies virus (Sauer, B., et al., Gene 70:331-341 (1988), herein incorporated by reference).

It has been found that certain E. coli enzymes inhibit efficient circularization of linear molecules which contain loxP sites at their termini. Hence, enhanced circularization efficiency can be obtained through the use of E. coli mutants which lack exonuclease V activity (Sauer, B., et al., Gene 70:331-341 (1988)).

Cre has been able to mediate loxP specific recombination in mammalian cells (Sauer, B., et al., Proc. Natl. Acad. Sci. (U.S.A.) 85:5166-5170 (1988), Sauer, B., et al., Nucleic Acids Res. 17:147-161 (1989), both references herein incorporated by reference.) Similarly, the recombination system has been capable of catalyzing recombination in plant cells (Dale, E.C., et al., Gene 91:79-85 (1990)).

## 2. The Flp Recombination System

Yeast express a recombinase known as "Flp" which catalyzes the site-specific inversion of a region of the yeast 2- $\mu$  circle plasmid (Schwartz, C.J. et al., J. Molec. Biol. 205:647-658 (1989); Parsons, R.L. et al., J. Biol. Chem. 265:4527-4533 (1990); Golic, K.G. et al., Cell 59:499-509 (1989); Amin, A.A. et al., J. Molec. Biol. 214:55-72 (1990)). The flp gene has been cloned, and the site ("FRT") which is recognized by the Flp recombinase has been determined (Vetter, D., et al., Proc. Natl. Acad. Sci. (U.S.A.) 80:7284-7288 (1983), herein incorporated by reference). The organization of the FRT sequence is similar to that of the loxP sequence recognized by Cre; however, the sequence contains a nearly perfect inverted repeat and a direct repeat. Flp-mediated recombination optimally occurs in vitro at a pH of between 6.6 and 8.0. A divalent cation such as  $Mg^{++}$  or spermidine is required. The Flp protein can recombine 2- $\mu$  derivatives having

-28-

directly repeated FRT sites to produce two circular plasmids. It does not require either host factors or supercoiled substrates (Vetter, D., et al., Proc. Natl. Acad. Sci. (U.S.A.) 80:7284-7288 (1983)).

5

### 3. The Gin/Fis Recombination System

The E. coli bacteriophage Mu has been found to encode a protein (Gin) which can mediate recombination at specific sites in the Mu genome (Mertens, G. et al., EMBO J. 3:2415-2421 (1984); Mertens, G. et al., J. Biol. Chem. 261:15668-15672 (1986)). The reaction causes a site-specific inversion of the Mu G segment and results in an altered host range. Recombination occurs at 34 base pair long sites, which must be arranged as inverted repeats on supercoiled DNA.

15

In order for inversion to occur, the phage encoded gin gene must be expressed. The product of this gene (Gin) binds to sites in each of the inverted repeats in a cooperative manner, induces a two base pair staggered nick, and forms a covalent linkage with DNA at the 5' end of each nick. In the presence of Gin alone, DNA inversion occurs with low frequency both in vitro and in vivo. In order to stimulate inversion, an E. coli host factor, known as "Fis" is typically required, unless a Fis-independent Gin mutant (i.e. a protein capable of catalyzing recombination in a site-specific manner without host factor) is employed. Such a mutant is disclosed by Klippel, A. et al. (EMBO J. 7:3983-3989 (1988)).

20

25

### 4. The $\lambda$ Recombination System

The site-specific recombination system of the E. coli bacteriophage  $\lambda$  has been well characterized (Weisberg, R. et al., In: Lambda II, (Hendrix, R. et al., Eds.), Cold Spring Harbor Press, Cold Spring Harbor, NY, pp. 211-250 (1983), herein incorporated by reference. Bacteriophage

30

-29-

λ uses this recombinational system in order to integrate its genome into that of its host, the bacterium E. coli. The system is also employed to excise the bacteriophage from the host genome in preparation for virus' lytic growth.

The recombination system is composed of four proteins- Int and Xis, which are encoded by the virus, and two host factors encoded by the E. coli. These proteins catalyze site-specific recombination between "Att" sites.

The λ Int protein (together with the E. coli host integration factors) will catalyze recombination between "AttP" and "AttB" sites. If the AttP sequence is present on a circular molecule, and the AttB site is present on a linear molecule, the result of the recombination is the disruption of both Att sites, and the insertion of the entire AttP-containing molecule into the AttB site of the second molecule. The newly formed linear molecule will contain an AttL and an AttR site at the termini of the inserted molecule.

Even in the presence of host factors, the λ Int enzyme, by itself, is unable to catalyze the excision of the inserted molecule. Thus, the reaction is unidirectional. If a second λ protein, the λ Xis protein, is added to the reaction, the reverse reaction can proceed, and a site-specific recombinational event will occur between the AttR and AttL sites to regenerate the initial molecules.

The nucleotide sequence of both the Int and Xis proteins are known, and both proteins have been purified to homogeneity. Both the integration and the excision reaction can be conducted in vitro. The nucleotide sequences of the four Att sites has also been determined (Weisberg, R. et al., In: Lambda II, (Hendrix, R. et al., Eds.), Cold Spring Harbor Press, Cold Spring Harbor, NY, pp. 211-250 (1983), which reference has been herein incorporated by reference).

## 5. Other Site-Specific Recombination Systems

Any of a large number of additional site-specific recombination systems can be used in accordance with the methods of the present invention.

5        Such systems are discussed by Echols, H. (J. Biol. Chem. 265:14697-14700 (1990)), de Villartay, J.P. (Nature 335:170-174 (1988); Craig, N.L. (Ann. Rev. Genet. 22:77-105 (1988)), Poyart-Salmeron, C. et al. (EMBO J. 8:2425-2433 (1989)), Hunger-Bertling, K. et al. (Molec. Cell. Biochem. 92:107-116 (1990)), and Cregg, J.M. (Molec. Gen. Genet. 219:320-323 (1989)), all herein incorporated by reference. Examples of preferred additional recombination systems include: the TpnI and the  $\beta$ -lactamase transposon systems (Levesque, R.C., J. Bacteriol. 172:3745-3757 (1990)); the Tn3 resolvase system (Flanagan, P.M. et al., J. Molec. Biol. 206:295-304 (1989); Stark, W.M. et al., Cell 58:779-790 (1989)); the yeast recombinase systems (Matsuzaki, H. et al., J. Bacteriol. 172:610-618 (1990)); the B. subtilis SpoIVC recombinase system (Sato, T. et al., J. Bacteriol. 172:1092-1098 (1990)); the Hin recombinase system (Glasgow, A.C. et al., J. Biol. Chem. 264:10072-10082 (1989)); the immunoglobulin recombinase systems (Malynn, B.A. et al., Cell 54:453-460 (1988)); the Cin recombinase system (Hafter, P. et al., EMBO J. 7:3991-3996 (1988); Hubner, P. et al., J. Molec. Biol. 205:493-500 (1989)); the Pin recombinase system (Plasterk, R.H.A. et al., Cold Spring Harbor Symp. Quant Biol. 49:295-300 (1984); all of the above references are herein incorporated by reference.

## 30        II. RECOMBINATION-FACILITATED SEQUENCE ANALYSIS

As indicated above, the multiplex sequencing method of G.M. Church et al. (Science 240:185-188 (1988)) requires the construction of a large number of vector libraries. In contrast, the present invention achieves



-31-

the goal of multiplex sequencing without the need to construct multiple gene libraries. In accordance with the present invention, DNA or cDNA from any desired source is obtained, and cloned into a cloning site of any of the well-known prokaryotic, eukaryotic, or shuttle vectors, modified to contain a recombinational site. Examples of suitable vectors are provided by Maniatis, T., et al. (In: Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Cold Spring Harbor, NY (1982)). The sequence information is then obtained through the use of a novel method.

Of particular importance to the present invention is the fact that recombination between two linear molecules results in the exchange of their DNA (Figure 4).

Thus, if in Figure 4, the sequence 1-7 represented a target molecule of unknown sequence, and the sequence H I contained a detectable marker, the result of the recombination would be to link the detectable marker to the target molecule.

Any of the above-described recombinases and their corresponding recombinational sites may be employed. As used herein, a "recombinational site" is a region of a DNA molecule of a sequence and size sufficient to permit it to function as a substrate in a recombinational reaction when provided with a suitable recombinase, and a second DNA molecule having a suitable recombinational site. Where the recombinase is a general recombinase, the recombinational site can be of any size or sequence. More preferably, however, the recombinational sites are selected so as to be capable of serving as a substrate in a recombinational reaction catalyzed by a site-specific recombinase, preferably Cre. For example, where the recombinational sites are loxP sites, the recombinase would be Cre; where the recombinational sites are attP and attB sites, the recombinase would be Int. Any other combination of recombinase and recombinational site may be

-32-

used. An example of the use of Cre and loxP sites is shown in Figure 5.

Recombination between a DNA molecule containing a loxP site and a double-stranded oligonucleotide containing a loxP site thus causes a "cut" and religation at the loxP sites. By using high molar ratios of such oligonucleotides to target molecules, the reactions may be driven toward completion.

The present invention exploits this capacity through the use of oligonucleotides which contain loxP sites. Such oligonucleotides may be of any length, however, it is preferable to employ oligonucleotides of 20-50 (and preferably about 34) base pairs. Such an oligonucleotide will contain a recombinational site, which may be at a terminus of the molecule, or may be flanked by other bases of the oligonucleotide.

Where one desires to use a vector to sequence a target molecule, the loxP site of the vector is preferably located as close as possible to the target sequence (thus minimizing the size of the vector sequence which would need to be sequenced in order to complete the desired sequencing of the target).

Where one desires to map the restriction sites of a target molecule, the loxP site is preferably located about 500 bases away from the target sequence (if a plasmid is employed) or 1000-2000 bases away from the target sequence (if a cosmid is employed).

As indicated above, the vectors of the present invention contain at least one "recombinational site." The loxP site is the preferred recombinational site of the present invention. In the most preferred embodiment, the vector shall contain one loxP site. The recombinational site will be incorporated into the vector at a location near the location of the cloning region. The structure of the vector is thus depicted in Figure 6 (where the recombinational site is illustrated as a loxP site; the

-33-

orientation of the loxP is always left to right relative to the other elements shown).

5 In an alternative preferred embodiment, the vector shall contain two recombinational sites, which shall flank the cloning region. Most desirably, the two recombinational sites shall differ from one another such that it is possible to mediate recombination at each recombinational site without mediating recombination at the other. This can be accomplished, for example through  
10 the use of vectors having a wild type and a mutant sites (such as a wild type and a mutant loxP site), or through the use of a vector having non-corresponding sites (such as a loxP site and an attP site, etc.). The orientation of the regions are such that the following structure is  
15 formed (illustrated using loxP and the loxP mutant site, loxP511) (Figure 7).

Having now generally described the invention, the same will be more readily understood through reference to the following examples which are provided by way of  
20 illustration, and are not intended to be limiting of the present invention, unless specified.

#### EXAMPLE 1 RESTRICTION FRAGMENT ANALYSIS

In accordance with this aspect of the present  
25 invention, a DNA molecule whose sequence is to be mapped by restriction endonuclease digestion is obtained from a suitable source. Preferably, such target DNA has been isolated using a restriction endonuclease which is also capable of cleaving at a site within the cloning region of  
30 the above-described vectors. Alternatively, conventional methods can be used to adapt the ends of the target such that they are now capable of being ligated into a restriction site of the cloning region. This can, for example, be accomplished with any target by treating  
35 overhanging ends to produce a blunt ended target molecule.

-34-

Once the ends of the target molecule have been so prepared, one of the above-described vector molecules is cleaved with a restriction endonuclease capable of cleaving the vector within the cloning region, and forming  
5 termini which are capable of ligating with the target molecule.

The target molecule is then introduced into the vector, and the vector is recircularized through the action of a DNA ligase. Procedures for accomplishing  
10 these steps are disclosed by Maniatis, T., et al. (In: Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Cold Spring Harbor, NY (1982)).

The use of a vector having two different recombinational sites facilitates the analysis of restriction sites  
15 at both ends of the target molecule. For the purposes of illustrating the invention, however, a vector containing a single recombinational site is depicted. The insertion of the target sequence is shown in Figure 8 and Figure 9. This vector is then incubated in the presence of Cre and  
20 a loxP-containing oligonucleotide having one or more probe primer regions (Figure 10). The resulting recombination creates a linear molecule, shown in Figure 11.

When such a molecule is subjected to partial restriction endonuclease digestion with a restriction  
25 enzymes, and analyzed by electrophoresis, a set of nested fragments containing the probe/primer region is obtained (Figure 12).

Significantly, since these nested fragments contain the probe/primer region of the original molecule, a probe  
30 having a sequence substantially complementary to that of the probe/primer region will be able to hybridize to the fragments. By labelling such a probe, it is thus possible to visualize the nested fragments which hybridize to the probe/primer region. Moreover, because the molecules  
35 share one end in common, the position of the restriction sites can be readily determined by measuring the sizes of the "bands," as shown in Figure 13.

-35-

It is possible to prepare multiple sets of nested fragments using different restriction endonucleases, and loxP-containing oligonucleotides having different probe/primer regions. Each set of nested fragments can be visualized by incubating the total set of fragments with a probe substantially complementary to the probe/primer region of the respective set of nested fragments. Thus, through the use of multiple sets of different probes (each capable of hybridizing to a different probe/primer region), the present invention permits one to sequentially analyze all of these nested sets of fragments. Indeed, if different labels are employed on different probes, it is possible to simultaneously analyze different sets of nested fragments.

By conducting the analysis using sets of loxP-containing oligonucleotides which contain two probe/primer regions, one common to all of the oligonucleotides, and one that is varied to permit the individual visualization of a single set of nested fragments, it is possible to visualize all of the sets of nested fragments.

Since such analyses may be conducted from a single gel, the present invention greatly facilitates the process of restriction mapping.

## EXAMPLE 2 DNA SEQUENCE ANALYSIS

In this aspect of the present invention, the sequence of a cloned region can be determined in a multiplex analysis.

The method utilizes two types of DNA molecules. The first molecule is a cloning vector which contains a loxP site. Any of the well-known prokaryotic, eukaryotic, or shuttle vectors may be modified to permit their use in the present invention. The vector shall contain at least one recombinational site, preferably loxP, which precedes and is adjacent to and, most preferably,

-36-

immediately adjacent to a cloning region. The structure of the vector is as shown in Figure 6 or Figure 7.

5 In accordance with the sequencing method of the present invention, a DNA molecule whose sequence is to be determined (i.e. a "target" sequence) is obtained from a suitable source. Preferably, such target DNA has been isolated using a restriction endonuclease which is also capable of cleaving at a site within the cloning region of the above-described vector. Alternatively, conventional  
10 methods can be used to adapt the ends of the target such that they are now capable of being ligated into a restriction site of the cloning region. This can, for example, be accomplished with any target by treating overhanging ends to produce a blunt ended target molecule.

15 Once the ends of the target molecule have been so prepared, one of the above-described vector molecules is cleaved with a restriction endonuclease capable of cleaving the vector within the cloning region, and forming termini which are capable of ligating with the target  
20 molecule.

The target molecule is then introduced into the vector, and the vector is recircularized through the action of a DNA ligase. Procedures for accomplishing these steps are disclosed by Maniatis, T., et al. (In:  
25 Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Cold Spring Harbor, NY (1982)). The construction is as shown in Figure 8 and Figure 9.

Once the target DNA has been inserted into any of the above-described vectors, it can then be amplified, by  
30 propagating the vector in a suitable host. Individual members of the library (either as transformed cells, or isolated DNA) can then be isolated.

In order to accomplish multiplex sequencing of the vector, one permits the vector to undergo site-specific  
35 recombination with a linear DNA molecule having a recombinational site, and at least one probe/primer region, located near the recombinational site (Figure 10).

-37-

Preferably, the linear molecule will have probe/primer regions on both sides of the recombinational site.

In the presence of a suitable recombinase (such as Cre), the vector and the linear molecule recombine to form  
5 a linear molecule as shown in Figure 14.

If the sequencing reaction is done using the Maxam-Gilbert method, then a probe capable of hybridizing to probe/primer I will identify one set of nested sequencing reaction products.

10 Alternatively, the target can be sequenced using the Sanger method by employing a primer having a 3'OH terminus which is capable of hybridizing to the probe/primer sequence of the fragment. Extension of this primer creates a set of nested sequencing reaction products.

15 Significantly, a probe/primer is capable of identifying only that nest of reaction products which has a probe/primer sequence to which it can hybridize. Thus, the use of a probe capable of hybridizing to a different probe/primer sequence will identify a different set of  
20 nested sequencing reaction products.

This feature of the present invention permits a multiplex analysis to be performed. To accomplish this, members of the unfractionated vector library are separately permitted to recombine with one of a plurality  
25 of linear molecules each of which differs from the other in the sequence of its probe/primer sequence. The result of such recombination may be depicted as shown in Figure 15.

Since a probe/primer capable of hybridizing to one  
30 probe/primer sequence (for example probe/primer sequence I in Figure 15) will identify only that set of nested sequence reaction products which contains the probe/primer sequence, all of the sequence reactions may be combined and analyzed on the same sequencing gel, by sequentially  
35 hybridizing with a different probe/primer sequence. Thus, where probe/primers capable of hybridizing to probe/primer sequences I, II, III, and IV are used, a single sequencing



-38-

gel can be used to determine the sequence of target molecules A, B, C and D.

Use of a vector having a second recombinational site (such as loxP511) adjacent to the second end of the inserted target molecule, in conjunction with a linear molecule (such as shown in Figure 16) permits one to identify a set of nested sequencing reaction products of the other strand of the target molecule. Thus, such a probe would permit the sequencing of the second strand of a DNA molecule (or equivalently, would yield sequence information relevant to the 3' end of the sequence of depicted target molecule A).

Significantly, the invention thus permits the sequencing of both strands of a DNA molecule on the same sequencing gel.

As indicated above, in one embodiment, the linear molecule will contain more than one adjacent probe/primer regions separated by a recombinational site (Figure 17). If one employs a set of such linear molecules in which probe/primer Z or W are kept invariant, whereas the sequence of probe/primer sequence I or II is varied, then one has the capacity to perform either a multiplex sequence analysis using probe/primers capable of hybridizing to the sequence of probe/primer sequence I or II or their variants or non-multiplex sequence analysis (using probe/primers capable of hybridizing to the sequence of probe/primer sequence Z or W).

### EXAMPLE 3 DETAILED DESCRIPTION OF MULTIPLEX SEQUENCE ANALYSIS

To perform multiplex sequence analysis, a series of oligonucleotides are constructed. These oligonucleotides will have a probe/primer region, which may have either of two general structures (depicted using loxP as the recombinational site): Oligonucleotide I (Primer #n - loxP) or Oligonucleotide II (Primer - Probe #n - loxP) where the



-39-

n indicates the number of different oligonucleotides in the series.

5 The above-described vectors are incubated in the presence of a recombinase and each of the oligonucleotides of the series, in separate reactions. To illustrate this aspect of the invention, using the loxP / Cre system, after Cre-mediated recombination with a plasmid of the type shown in Figure 8 and with an Oligonucleotide I (Primer #n - loxP), the linear molecule shown in Figure 10 18A would be produced.

After Cre-mediated recombination with a plasmid of the type shown in Figure 8 with an Oligonucleotide II (Primer - Probe #n - loxP) the linear molecule shown in Figure 18B would be produced.

15 As will be appreciated, either of such molecules can be employed to determine the sequence of the target molecule using either the Sanger or Maxam-Gilbert methods. Where oligonucleotides of class I are employed, n different primers would be added to each sequencing reaction, and the probes to detect the sequencing products 20 would be the complements of the primers. Since such different probes are being used, it is possible to analyze all n sequence reactions using a sequencing gel, through a multiplex sequence analysis. Where oligonucleotides of class II are employed, a single primer would be used in 25 the sequencing reactions, and different probes (1, 2, etc.) would be used. This latter embodiment is preferred, except that target sequence would not be reached until about 77 nucleotides from the 5' end of the primer (i.e. 20 nucleotides of the primer, 20 nucleotides of the probe, 30 34 nucleotides of the loxP site, and 3 nucleotides from the remainder of the cloning site (e.g. SmaI). When one wishes to eliminate the need to sequence 20 of these nucleotides, one would include a deoxyuracil (dU) toward 35 the 3' end of the primer, and treat with the enzyme UDG (Uracil DNA Glycosylase) just before running the sequencing gel. This treatment renders the sites abasic,

-40-

but does not cleave the phosphodiester backbone of the DNA molecule. Cleavage may be accomplished by heating the reaction, or by incubating it in the presence of an enzyme (such as endonuclease IV of E. coli) capable of specifically cleaving nucleic acid molecules at abasic sites. Note that after Cre-loxP recombination, the priming sites would be completely single stranded, even without denaturation, since the recombination oligonucleotide would be single stranded in the primer domain.

The above method permits one to sequence from one side of a target molecule. The use of a vector containing two recombinational sites permits one to sequence from both sides of the molecule.

Table 1 compares the ability of the vectors and methods of the present invention, with those of the multiplex method, to facilitate multiplex sequence analysis of ten sequencing reactions with a target DNA molecule.

-41-

TABLE 1				
ATTRIBUTE BEING COMPARED	Cre/ <u>loxP</u> System		Multiplex Method	
	Oligonucleotide			
	I	II		
5	Number of Vectors Required	1	1	5
	Number of Libraries Required	1	1	5
10	Number of Recombination Oligonucleotides Required	10	10	0
	Number of Primers Required	10	1-2	2
15	Approximate Number of Nucleotides from 5' End of Primer to Target	57	77	43
	Number of Probe Oligonucleotides Required	10	10	10
20	Recombinase Required (Cre)	Yes	Yes	No

The present invention includes articles of manufacture, such as "kits." In one embodiment, such kits will, typically, be specially adapted to contain in close compartmentalization a first container which contains a DNA vector, which has at least one recombinational site (I); a second container which contains at least one probe/primer DNA molecule having a recombinational site (II), and a probe/primer region; and a recombinase capable of mediating recombination between site (I) of the DNA vector and site (II) of the probe/primer DNA molecule. The kit may additionally contain multiple probe/primer DNA molecules, which may be used to facilitate the multiplex sequence analysis of DNA in accordance with the methods of the invention. The kit may additionally contain instructional brochures, and the

-42-

like. It may also contain reagents sufficient to accomplish DNA sequencing.

5 In a second embodiment, such kits will, typically, be specially adapted to contain in close compartmentalization a first container which contains a DNA molecule (such as a linear oligonucleotide or a vector) which has at least one recombinational site (I); a second container which contains at least one DNA molecule having a recombinational site (II), and is detectably labelled; and  
10 a recombinase capable of mediating recombination between the recombinational site (I) of the DNA molecule and site (II) of the labelled oligonucleotide. The kit may additionally contain instructional brochures, and the like. It may also contain reagents sufficient to  
15 accomplish DNA sequencing.

#### EXAMPLE 4 SEQUENCING OF A COSMID MOLECULE

The present invention facilitates the sequencing of cosmid molecules. In this method, a cosmid is constructed  
20 so as to contain a loxP site (Figure 19). The molecule is incubated in the presence of Cre and a loxP-containing oligonucleotide, preferably, the oligonucleotide is single-stranded, and will possess a sequence which causes it to snap back upon itself (Figure 20). As a result of  
25 such incubation, a linear molecule will be produced having the structure shown in Figure 21. Upon restriction endonuclease digestion, an array of partial-digestion products such as those shown in Figure 22 are obtained.

As will be recognized, the effect of the reaction has  
30 been to produce a series of oligonucleotides which contain at most, only one loxP site. This mixture of oligonucleotides is then incubated with a DNA ligase in the presence of a second loxP-containing oligonucleotide, which will preferably be single-stranded, and possess a  
35 sequence which causes it to snap back upon itself, such as

-43-

shown in Figure 20. As a result of such incubation, three general classes of molecules will be present in the reaction:

- (I) those with only one loxP site,
- 5 (II) those having two loxP sites in a direct repeat (Figure 23), and
- (III) those having two loxP sites in an inverted repeat (Figure 24).

10 As will be perceived, only molecules which contain the target sequences that were initially bound to the loxP site of the first DNA molecule (i.e. A-B sequences) will contain two directly repeated loxP sites (i.e. class II molecules). Such molecules thus contain target DNA rather than cosmid vector DNA.

15 The mixture of molecules will then preferably be separated, as with agarose gel electrophoresis, or other conventional means, and the different sizes of molecules eluted or otherwise recovered.

20 The directly repeat loxP sites present on these molecule permits one, in a Cre-mediated reaction, to recombine the cloned DNA between these sites into any of the loxP-containing vectors discussed above.

25 Thus, this method permits one to subclone target DNA from a cosmid into a smaller vector. Significantly, the cloned DNA is manipulated such that it becomes flanked with directly repeating loxP sites. Moreover, the method permits one to obtain and clone a set of nested oligonucleotide fragments of a desired target molecule.

#### 30 EXAMPLE 5 MULTIPLEX RESTRICTION FRAGMENT ANALYSIS

The use of Cre/loxP mediated site-specific recombination as a method to facilitate multiplex mapping was demonstrated by the following procedure. The target molecules were pLox, a 2.9 kb plasmid with a loxP site  
35 cloned into a polylinker region, and pSPORT-lox, a 4.1 kb

-44-

plasmid with a loxP site inserted into its multiple cloning site (MCS).

5 The sequences chosen for hybridization probes were taken from Church et al. (Science 240:185-188 (1988)) and the hybridization, washing, and probe stripping procedures disclosed therein were used with minor modification. Specifically, the hybridization probes used were (in the Church et al. nomenclature) PO1, PO2, PO3 and PO4.

10 Recombinant molecules to be eventually used as substrates for multiplex mapping were generated as follows. A partially duplex molecule which was composed of the oligonucleotides, as shown in Figure 25, was incubated with pSPORT-lox in the presence of Cre under conditions sufficient to permit recombination to occur. 15 Specifically, the reaction contained 1 pmol of plasmid, 4 pmol of oligonucleotides and 5 units of Cre (NEN) in buffer composed of 50 mM Tris-HCl (pH 7.5), 33 mM NaCl, 5 mM spermidine, 0.5 mg/ml bovine serum albumin (BSA); incubations were at 37°C for 15 minutes. A separate 20 reaction containing pLox and a second partially duplex oligonucleotide of the structure shown in Figure 26 was also incubated in the presence of Cre such that recombination took place. After inactivation of the Cre by heating to 65°C for 10 minutes, portions of these 25 recombination reactions were either kept separate or mixed together and then subjected to partial digestion with the restriction endonuclease HaeIII or HhaI. The products were resolved on an agarose gel. After electrophoresis, an overnight alkaline transfer to a charged nylon membrane 30 (BioDyne-B) was performed (Reed and Mann, Nucleic Acids Res. 13:7207-7221 (1985)).

Pre-hybridizations and hybridizations were performed in the buffers of Church et al. (Science 240:185-188 (1988)), however, incubations were carried out at 37°C 35 rather than 42°C and hybridizations were extended overnight. Oligonucleotide probe (i.e., PO1, PO2, PO3 and PO4 [of Church et al., Science 240:185-188 (1988)]) were

-45-

labeled with T4 polynucleotide kinase and [ $\gamma^{32}\text{P}$ ] ATP. All washes were performed at room temperature and consisted of two washes with 6xSSC for 2 minutes each followed by washing with 2xSSC + 0.1% sodium dodecyl sulfate (SDS),  
5 2.5 minutes total with one change of buffer. (20xSSC = 3M NaCl, 0.3N Na Citrate, pH 7.0). Membranes were then subjected to autoradiography to determine the linear map of the respective restriction sites.

Probe was then stripped from the membrane by  
10 incubation in 2mM Na<sub>2</sub> EDTA + 0.1% SDS (adjusted to pH8.3 with Tris base) at 65°C for 10 minutes. Removal of probe was verified by autoradiography and the hybridization, washing and visualization process was repeated with a different radioactive probe.

15 This method was shown to be highly specific with no background from cross-hybridization. It yielded accurate fine structure maps of both substrates. Significantly, even within lanes which contained mixtures of the two targets, each pattern could be detected independently,  
20 sequentially, and with complete specificity.

While the invention has been described in connection with specific embodiments thereof, it will be understood that it is capable of further modifications and this application is intended to cover any variations, uses, or  
25 adaptations of the invention following, in general, the principles of the invention and including such departures from the present disclosure as come within known or customary practice within the art to which the invention pertains and as may be applied to the essential features  
30 hereinbefore set forth and as follows in the scope of the appended claims.

WHAT IS CLAIMED IS:

1. A method for analyzing a target DNA molecule, which comprises:

5 (A) forming a recombinant molecule, said recombinant molecule comprising a probe/primer sequence linked to a recombinational site (I), wherein said site is linked to the sequence of said target molecule; and

10 (B) analyzing said target molecule using a nucleic acid molecule capable of hybridizing to said probe/primer sequence, or its complement.

2. The method of claim 1, wherein said analysis comprises determining a nucleotide sequence of said target DNA molecule, wherein, in step (B), said analysis comprises determining the sequence of the target molecule using a nucleic acid molecule capable of hybridizing to said probe/primer sequence, or its complement.

3. The method of claim 2, wherein said recombinant molecule is formed by:

20 (1) introducing said target DNA molecule into at least one vector, having a recombinational site (II), to thereby form a vector-target DNA construct;

25 (2) incubating said vector-DNA construct in the presence of a recombinase, and a DNA molecule having said recombinational site (I) and said probe/primer region; wherein said incubation is under conditions sufficient to permit said recombinase to mediate recombination between said recombinational site (II) of said vector-target DNA construct and said recombinational site (I) of said DNA molecule; and

30 (3) permitting said recombinase to mediate recombination between said recombinational sites, and to thereby form said sequencing molecule.



-47-

4. The method of claim 3, wherein at least two vector-target DNA constructs are formed, and wherein at least two different DNA molecules each having a recombinational site (I) and further having a different probe/primer region are employed; and wherein said determining of the sequence of the target molecule is through use of two probes, each capable of hybridizing to only one of said probe/primer regions, or its complement.

5. The method of claim 3, wherein said recombination is site-specific recombination.

6. The method of claim 5, wherein in said site-specific recombination, said recombinase is Cre, and at least one of said recombinational sites (I) or (II) are loxP sites.

7. The method of claim 6, wherein said recombinational site (I) is a loxP site.

8. The method of claim 6, wherein said vector contains one wild-type loxP site and one mutant loxP site.

9. The method of claim 1, wherein said analysis comprises ordering restriction endonuclease recognition sites in a target DNA molecule, and wherein, in step (B), said analysis comprises

(i) incubating said recombinant molecule in the presence of a restriction endonuclease under conditions sufficient to permit said endonuclease to cleave DNA containing a cleavage site recognized by said endonuclease; and

(ii) determining the order of any restriction sites in said target molecule using a nucleic acid molecule capable of hybridizing to said probe/primer sequence, or its complement.

-48-

10. The method of claim 9, wherein said recombinant molecule is formed by:

5 (1) introducing said target DNA molecule into at least one vector, having a recombinational site (II), to thereby form a vector-target DNA construct;

10 (2) incubating said vector-DNA construct in the presence of a recombinase, and a DNA molecule having said recombinational site (I) and said probe/primer region; wherein said incubation is under conditions sufficient to permit said recombinase to mediate recombination between said recombinational site (I) of said DNA molecule, and said recombinational site (II) of said vector-target DNA construct; and

15 (3) permitting said recombinase to mediate recombination between said recombinational sites (I) and (II), and to thereby form said recombinant molecule.

20 11. The method of claim 10, wherein at least two vector-target DNA constructs are formed, and wherein at least two different DNA molecules each having a recombinational site (I) and further having a different probe/primer region are employed; and wherein said ordering of restriction sites of the target molecule is through use of two probes, each capable of hybridizing to only one of said probe/primer regions, or its complement.

25 12. The method of claim 10, wherein said recombination is site-specific recombination.

30 13. The method of claim 12, wherein in said site-specific recombination, said recombinase is Cre, and at least one of said recombinational sites (I) and (II) is a loxP site.

14. The method of claim 13, wherein said recombinational site (I) is a loxP site.

-49-

15. The method of claim 13, wherein said vector contains one loxP site and one mutant loxP site.

16. A kit specially adapted to mediate recombination between a DNA molecule having a recombinational site (I), and a DNA vector, having a recombinational site (II), said kit comprising in close compartmentalization:

- 1) a first container containing a recombinase capable of mediating said recombination between said site (I) of said DNA molecule and said site (II) of said vector; and
- 2) a second container containing a DNA molecule having said recombinational site (I).

17. The kit of claim 16, wherein said recombinase is Cre, and wherein at least one of said recombinational sites (I) and (II) is a loxP site.

18. The kit of claim 16, which additionally contains a third container containing said DNA vector.

19. The kit of claim 18, wherein said vector contains one loxP site.

20. The kit of claim 18, wherein said vector contains one wild-type loxP site and one mutant loxP site.

21. A set of nested oligonucleotides each of which has a first region of unknown sequence, and a second region of known sequence, wherein said second region comprises both a recombinational site, and a probe/primer region.

22. The set of oligonucleotides of claim 21, wherein at least one of said oligonucleotides is hybridized to a probe.

-50-

23. The set of oligonucleotides of claim 21, wherein at least one of said oligonucleotides is hybridized to a primer.

24. The set of oligonucleotides of claim 21, wherein  
5 said recombinational site is a loxP site.

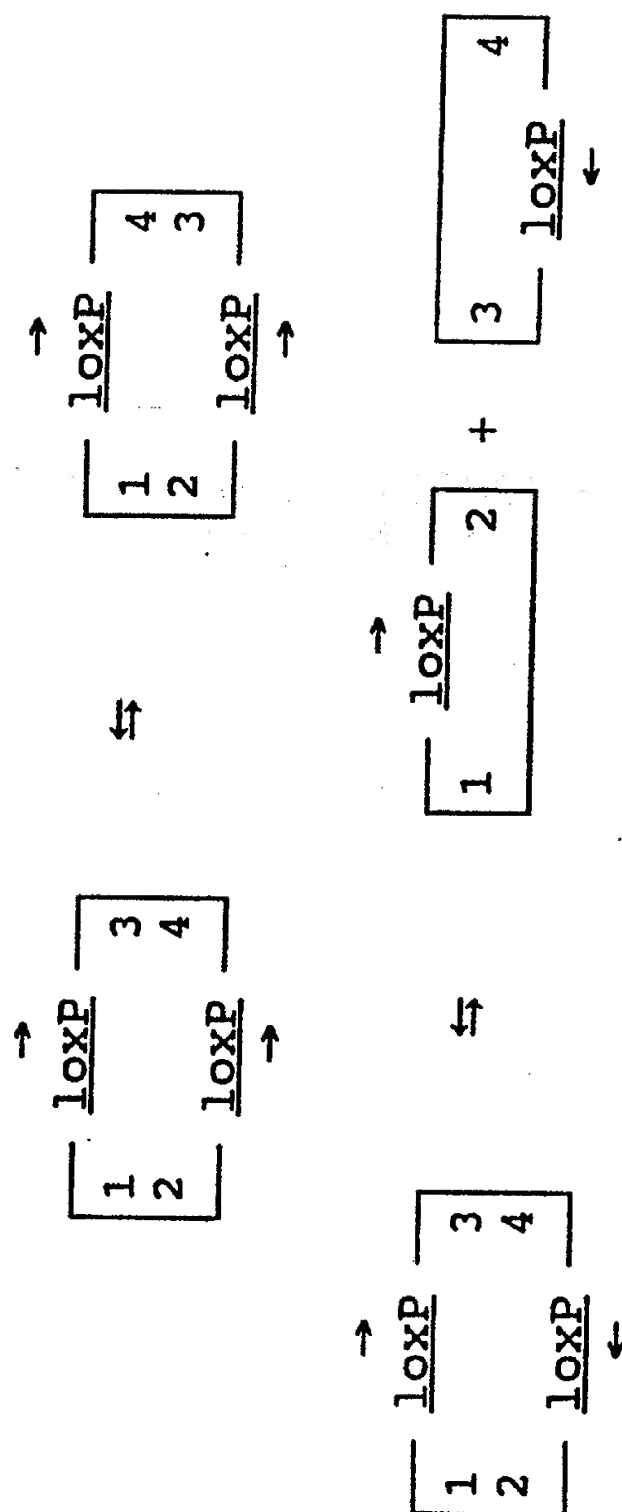


FIG. 1

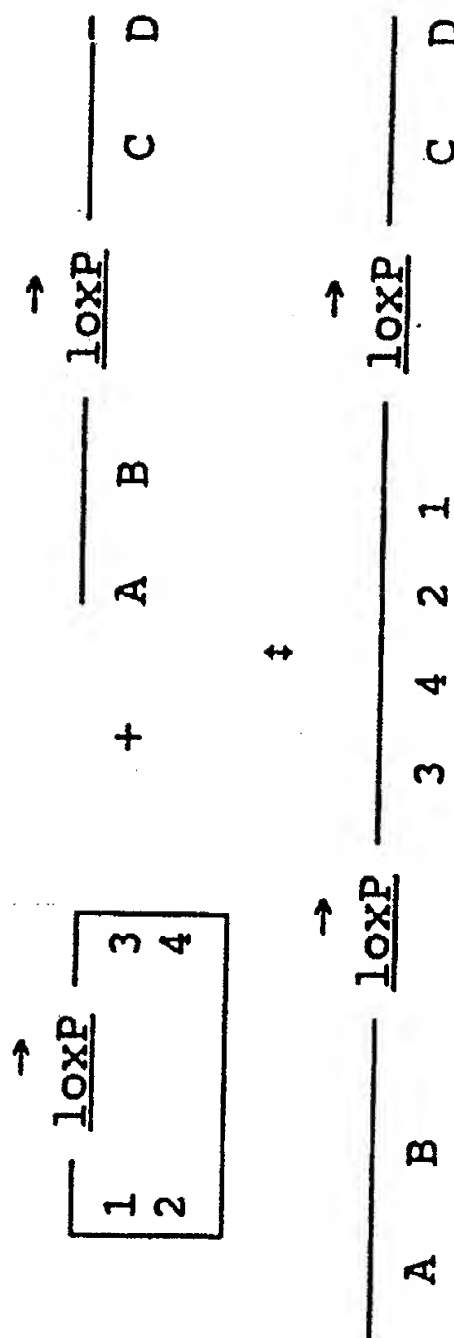


FIG. 2

2/10

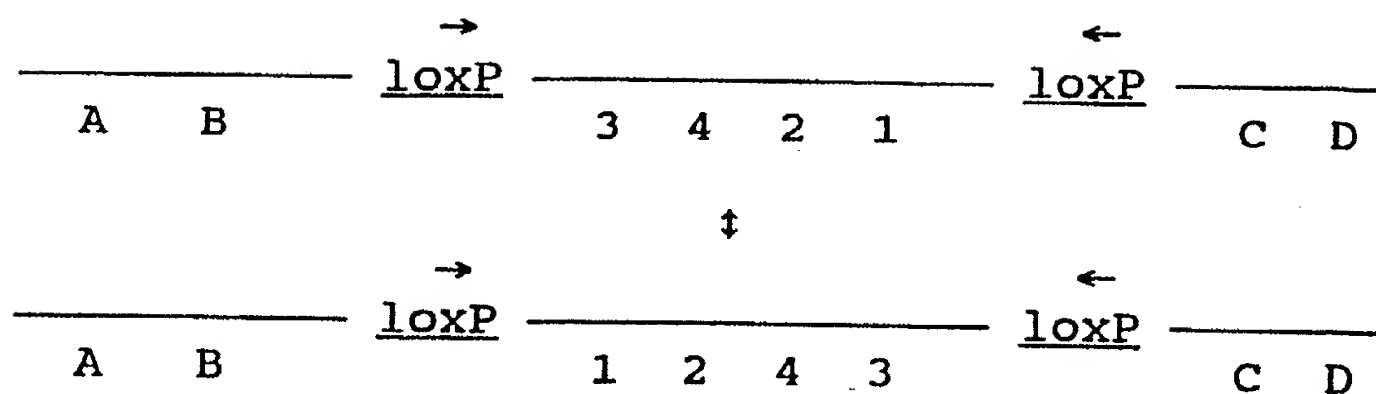


FIG. 3

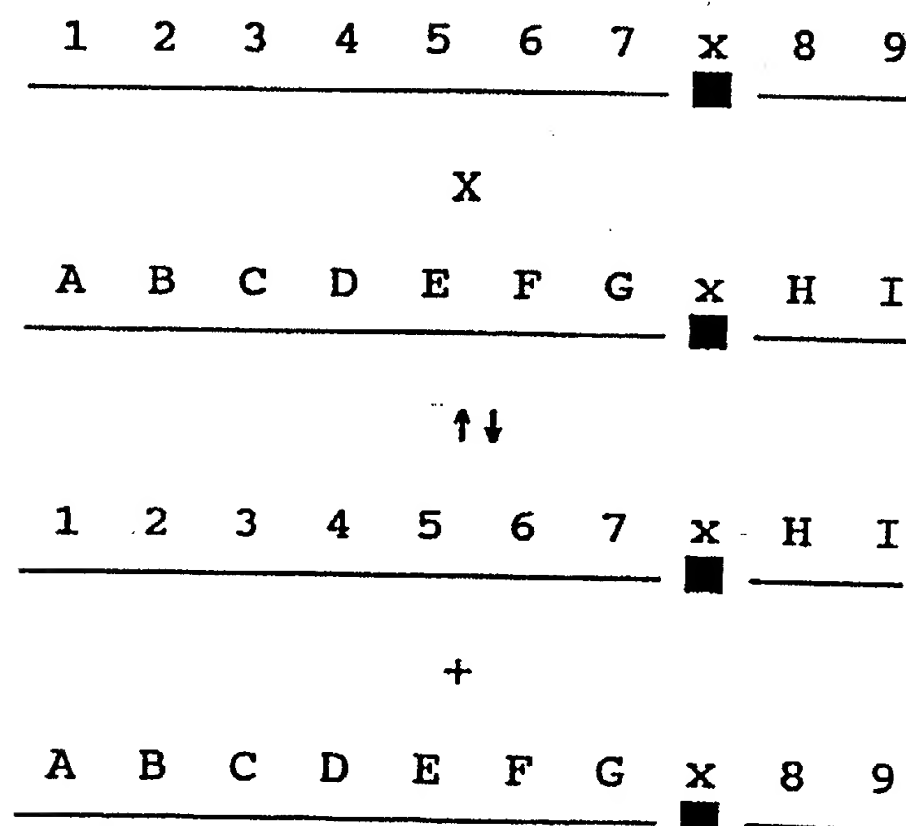


FIG. 4

3/10

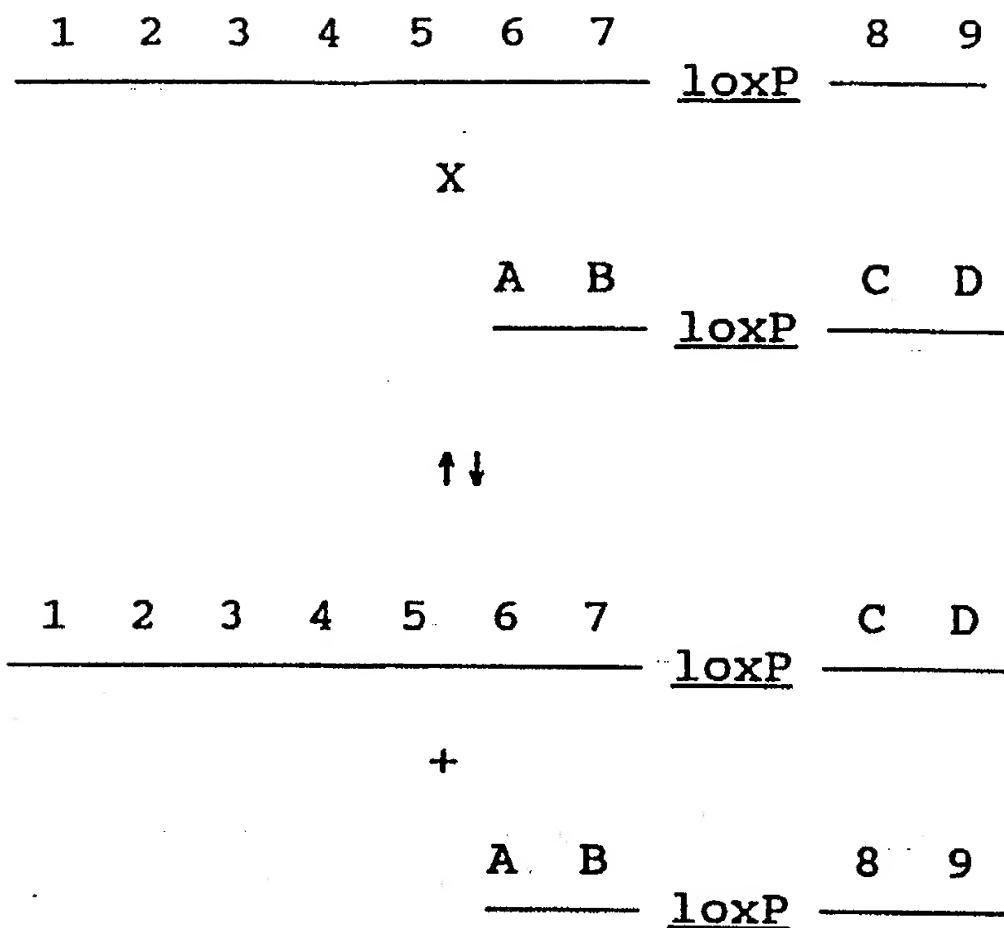


FIG. 5

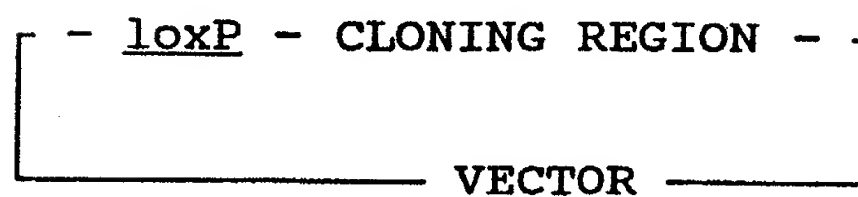


FIG. 6

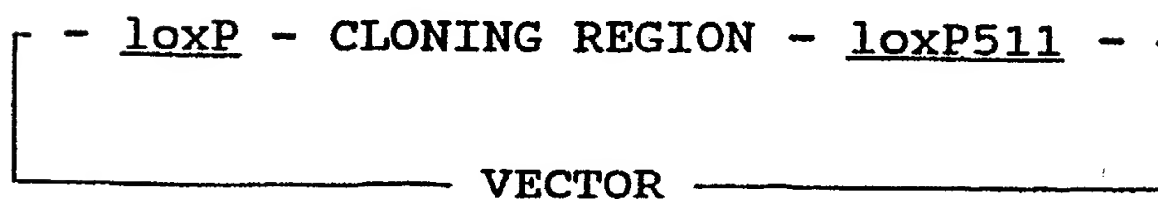


FIG. 7

4 / 10

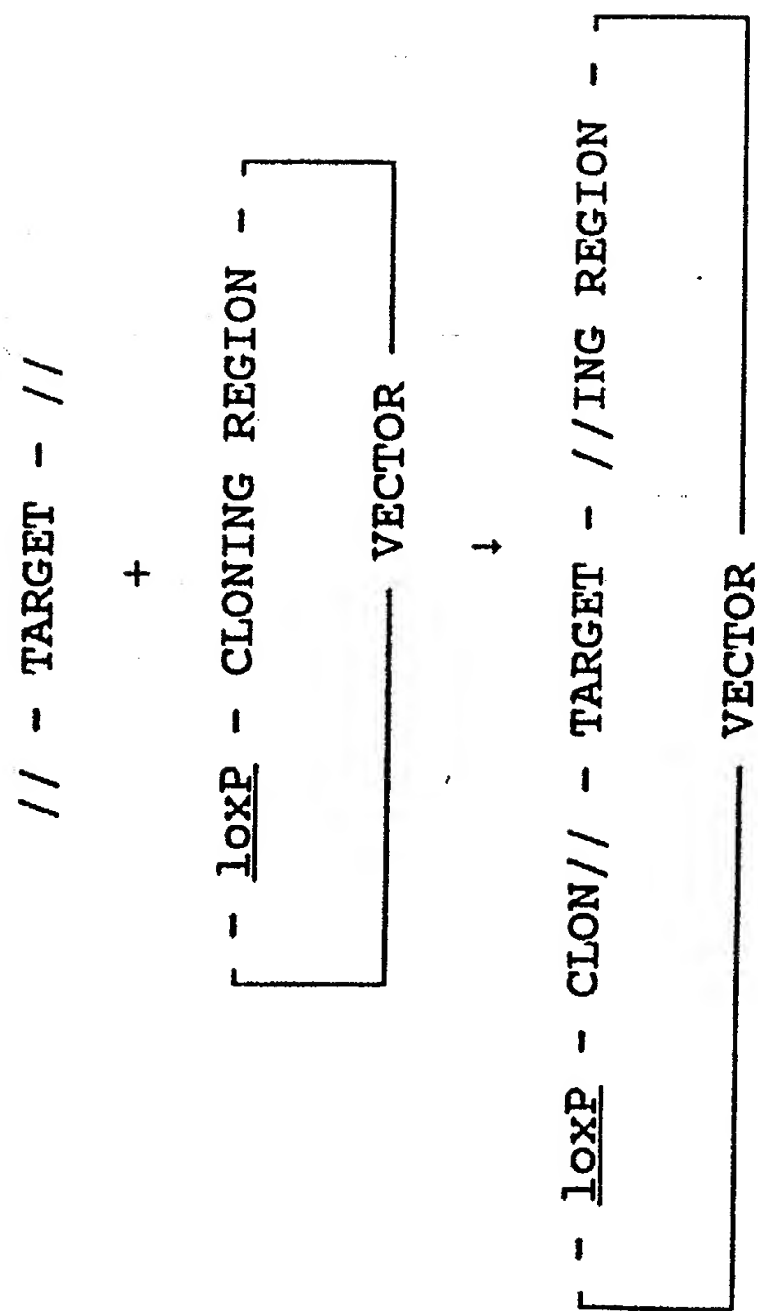
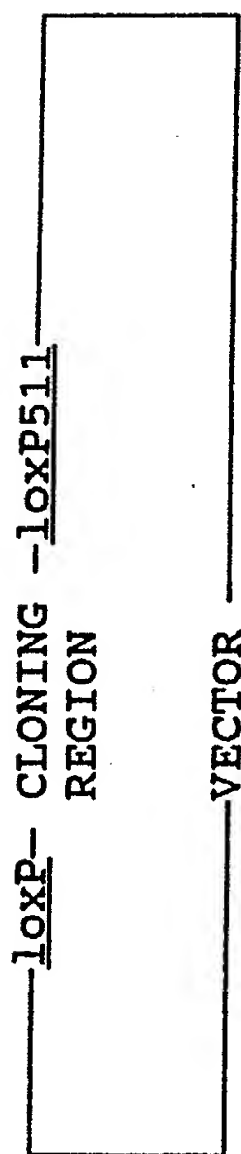


FIG. 8





+

// - TARGET - //

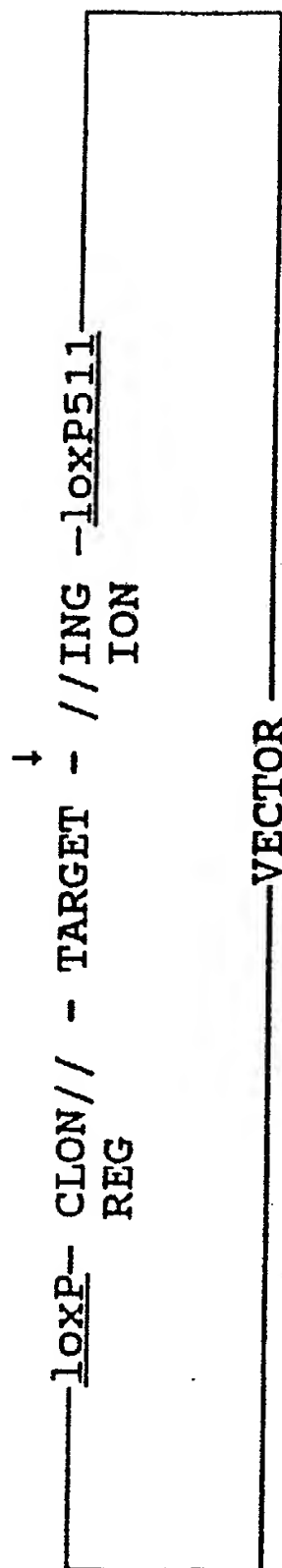


FIG. 9

probe/primer I ———loxP—  
 probe/primer Z ———probe/primer I ———loxP—

FIG. 10

probe/primer I —loxP-CLON// -TARGET-//ING —VECTOR—loxP—

FIG. 11

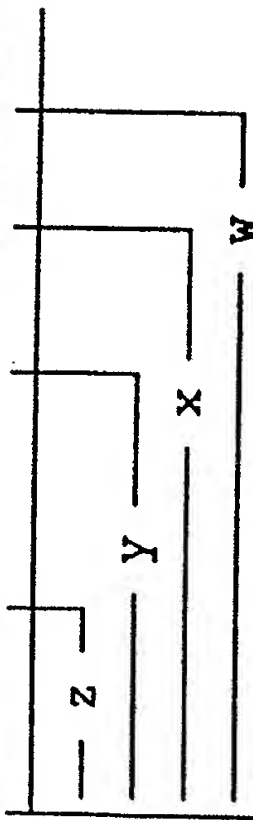
probe/primer I	—	<u>loxP</u>	—	CLON//	—	TARGET	—
probe/primer I	—	<u>loxP</u>	—	CLON//	—	TARGE	
probe/primer I	—	<u>loxP</u>	—	CLON//	—	TAR	
probe/primer I	—	<u>loxP</u>	—	CLON//	—	T	

FIG. 12

Size of  
bands

probe/primer I	—	<u>loxP</u>	—	CLON//	—	TARGET	—	w
probe/primer I	—	<u>loxP</u>	—	CLON//	—	TARGE		x
probe/primer I	—	<u>loxP</u>	—	CLON//	—	TAR		y
probe/primer I	—	<u>loxP</u>	—	CLON//	—	T		z

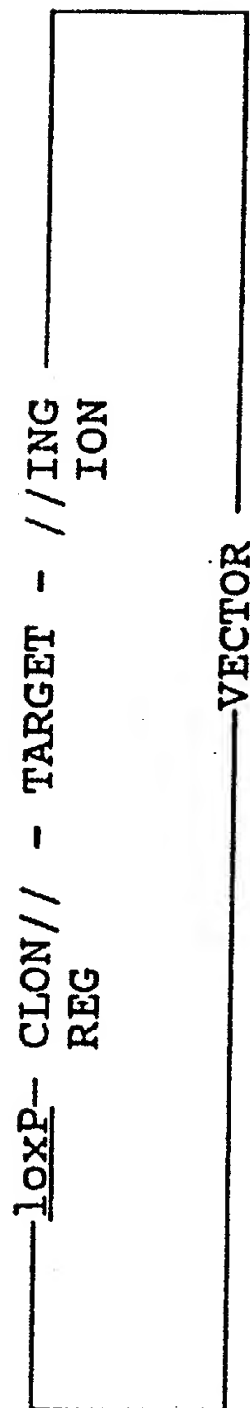
T	A	R	GE	T
---	---	---	----	---



Restriction sites in  
Target Molecule

FIG. 13

7/10



+

probe/primer I ——— loxp ———

↓

probe/primer I — loxp-CLON// - TARGET - //ING — VECTOR — loxp —  
REG ION

FIG. 14

probe/primer I — loxp-CLON// - TARGET A - //ING-VECTOR-loxp —  
probe/primer II — loxp-CLON// - TARGET B - //ING-VECTOR-loxp —  
probe/primer III — loxp-CLON// - TARGET C - //ING-VECTOR-loxp —  
probe/primer IV — loxp-CLON// - TARGET D - //ING-VECTOR-loxp —  
etc.

FIG. 15

probe/primer I ——— loxp511 ———

probe/primer Z ——— probe/primer I ——— loxp511 ———

FIG. 16

probe/ Z - probe/ I - loxP - probe/ II - probe/ W  
 primer primer primer primer

FIG. 17

Primer#n -loxP- CLON// -TARGET-//ING REGION - VECTOR - LoxP

FIG. 18A

Primer-Probe#n-loxP-CLON// -TARGET-//ING REGION-VECTOR-LoxP

FIG. 18B

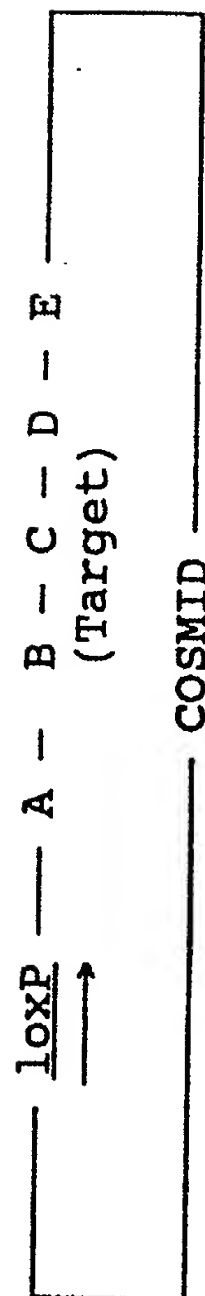


FIG. 19

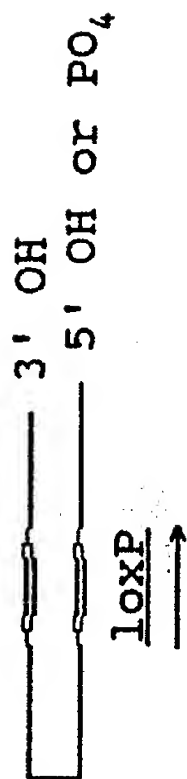


FIG. 20

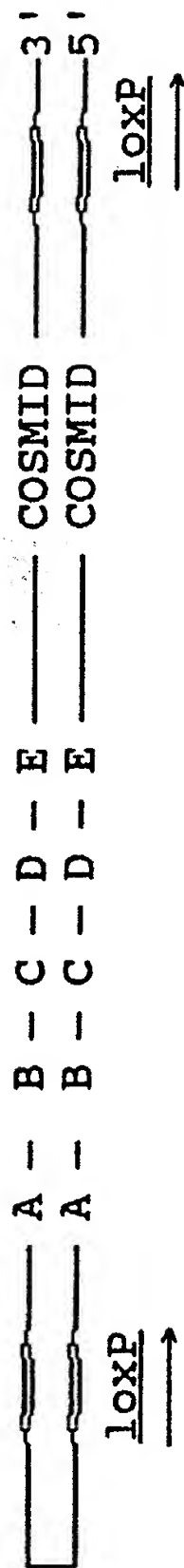
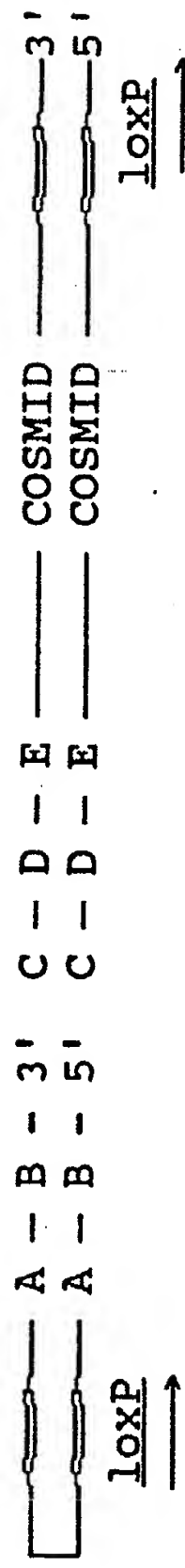
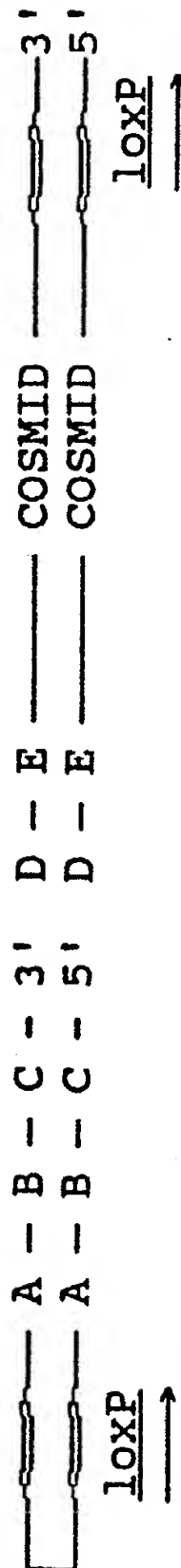


FIG. 21



etc.

FIG. 22

10/10

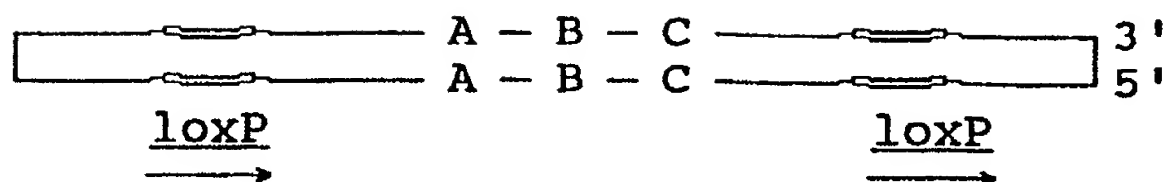


FIG. 23

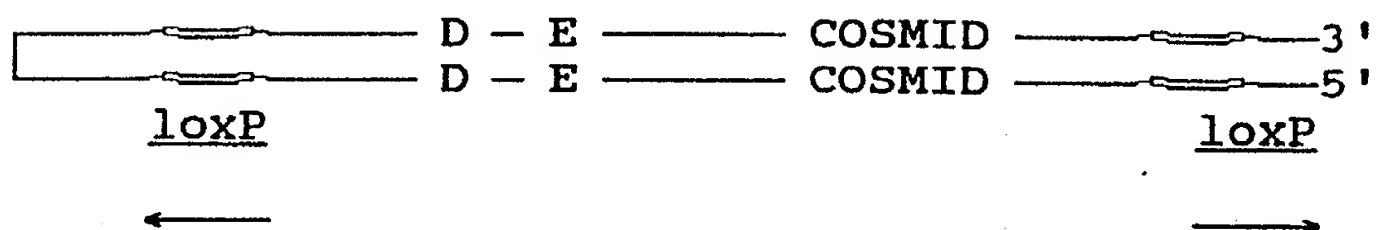


FIG. 24

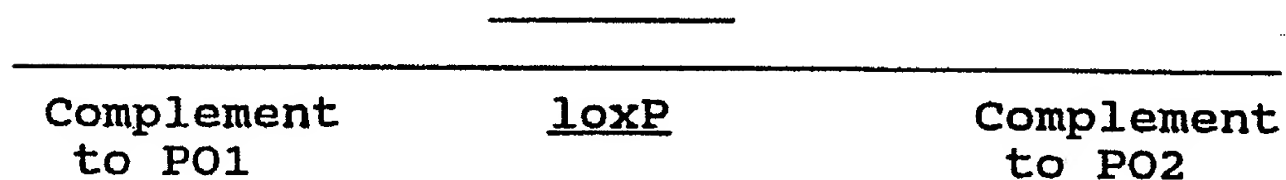


FIG. 25

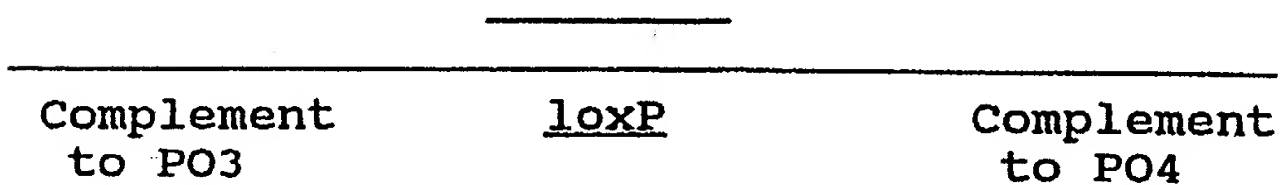


FIG. 26

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US92/04923

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(5) : C12N 15/10; C07K 13/00

US CL : 435/172.3; 530/358; 536/27

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/172.3; 530/358; 536/27

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched  
CHEMICAL ABSTRACTSElectronic data base consulted during the international search (name of data base and, where practicable, search terms used)  
APS

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US, A, 4,959,317 (Sauer) 25 September 1990, entire document.	1-24
Y	US, A, 4,942,124 (Church) 17 July 1990, see especially abstract.	1-24
Y	Nucleic Acids Research, Volume 14, No. 5, issued 1986, R. H. Hoess et al., "The role of the <u>loxP</u> spacer region in P1 site-specific recombination", pages 2287-2300, especially Table 1.	1-24



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be part of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier document published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Z" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

20 July 1992

Date of mailing of the international search report

20 AUG 1992

Name and mailing address of the ISA/  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Facsimile No. NOT APPLICABLE

Authorized officer

JAMES KETTER

Telephone No. (703) 308-0196